GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 2, 2004, 15:00:02 ; Search time 13693 Seconds
(without alignments)
11357.249 Million cell updates/sec

Title: US-10-780-347-1
Perfect score: 3588
Sequence: 1 gtcgacttattgcattgatg.....gatnaaattgagccgcttaa 3588

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	108.8	3.0	302100	1	AP005087	AP005087 Vibrio pa
c	2	105	2.9	248850	1	AP005348	AP005348 Vibrio vu
c	3	105	2.9	300883	1	AE016809	AE016809 Vibrio vu
c	4	99.4	2.8	300395	1	AE016785	AE016785 Pseudomon
c	5	91.6	2.6	301457	1	AE016924	AE016924 Chromobac
	6	82.8	2.3	190050	1	AL646080	AL646080 Ralstonia
	7	80	2.2	33154	1	AF361470	AF361470 Rhizobium
	8	79.4	2.2	14867	3	AE001398	AE001398 Plasmodiu
	9	76.8	2.1	252394	3	AE014833	AE014833 Plasmodiu
c	10	76.2	2.1	179902	2	AP003630	AP003630 Oryza sat
c	11	75.4	2.1	143409	9	AL162497	AL162497 Human DNA
	12	74.8	2.1	164188	2	BX005205	BX005205 Danio rer
	13	74.6	2.1	110000	2	PFMAL7P1_07	Continuation (8 of
	14	74.6	2.1	170627	2	AC125567	AC125567 Rattus no
c	15	74.6	2.1	182518	9	AC131649	AC131649 Homo sapi
	16	73.2	2.0	166885	9	AC098691	AC098691 Homo sapi
c	17	73	2.0	83703	5	AL732635	AL732635 Zebrafish
	18	72.8	2.0	11422	6	AX345121	AX345121 Sequence
	19	72.8	2.0	11422	6	AX348323	AX348323 Sequence
c	20	72.8	2.0	223021	2	AC119799	AC119799 Pan trogl
	21	72.8	2.0	349980	6	AX344564	AX344564 Sequence
c	22	72.6	2.0	5727	1	AF037441	AF037441 Edwardsie
	23	72.4	2.0	8093	6	AX598978	AX598978 Sequence
	24	72.4	2.0	8093	6	AX767538	AX767538 Sequence
	25	72.4	2.0	8093	6	AX795912	AX795912 Sequence
	26	72.4	2.0	8093	6	AX822432	AX822432 Sequence
	27	72.4	2.0	8093	6	AX826072	AX826072 Sequence
c	28	71.8	2.0	166228	5	AL929341	AL929341 Zebrafish
	29	71.4	2.0	61020	6	AX251546	AX251546 Sequence
	30	71.4	2.0	131033	5	AL591370	AL591370 Zebrafish
c	31	71.2	2.0	150865	5	AL929188	AL929188 Zebrafish
c	32	71.2	2.0	208140	2	BX255953	BX255953 Danio rer
c	33	70.8	2.0	13568	1	AE009363	AE009363 Agrobacte
	34	70.8	2.0	19961	1	AE008251	AE008251 Agrobacte
	35	70.8	2.0	110000	2	PFMAL6P1_02	Continuation (3 of
c	36	70.8	2.0	175026	5	AL954861	AL954861 Zebrafish
c	37	70.8	2.0	179435	2	AC134381	AC134381 Danio rer
	38	70.8	2.0	229480	5	AL929150	AL929150 Zebrafish
c	39	70.6	2.0	10029	1	AE013970	AE013970 Yersinia
	40	70.6	2.0	72243	9	AL731858	AL731858 Human DNA
c	41	70.6	2.0	194869	5	AL954136	AL954136 Zebrafish
	42	70.6	2.0	201050	1	AJ414143	AJ414143 Yersinia
	43	70.2	2.0	101034	9	AC096550	AC096550 Homo sapi
c	44	70.2	2.0	258658	3	AE014832	AE014832 Plasmodiu
	45	69.8	1.9	8666	6	AX767525	AX767525 Sequence

ALIGNMENTS

RESULT 1
AP005087/c
LOCUS AP005087 302100 bp DNA linear BCT 05-MAR-2003
DEFINITION Vibrio parahaemolyticus DNA, chromosome 2, complete sequence, 4/6.
ACCESSION AP005087 BA000032
VERSION AP005087.1 GI:28809161
KEYWORDS .
SOURCE Vibrio parahaemolyticus
ORGANISM Vibrio parahaemolyticus
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrio.
REFERENCE 1
AUTHORS Nasu,H., Iida,T., Sugahara,T., Yamaichi,Y., Park,K.S., Yokoyama,K.,
Makino,K., Shinagawa,H. and Honda,T.
TITLE A filamentous phage associated with recent pandemic Vibrio

parahaemolyticus O3:K6 strains
JOURNAL J. Clin. Microbiol. 38 (6), 2156-2161 (2000)
MEDLINE 20295086
PUBMED 10834969
REFERENCE 2
AUTHORS Makino,K., Oshima,K., Kurokawa,K., Yokoyama,K., Uda,T., Tagomori,K., Iijima,Y., Najima,M., Nakano,M., Yamashita,A., Kubota,Y., Kimura,S., Yasunaga,T., Honda,T., Shinagawa,H., Hattori,M. and Iida,T.
TITLE Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism distinct from that of V cholerae
JOURNAL Lancet 361 (9359), 743-749 (2003)
MEDLINE 22508454
PUBMED 12620739
REFERENCE 3 (bases 1 to 302100)
AUTHORS Oshima,K., Kurokawa,K., Makino,K., Yokoyama,K., Yasunaga,T., Honda,T., Shinagawa,H., Hattori,M. and Iida,T.
TITLE Direct Submission
JOURNAL Submitted (09-APR-2002) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:ken@gen-info.osaka-u.ac.jp, URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, Fax:81-6-6879-2047) .
COMMENT genome project
This clone was isolated from a patient presenting with acute gastroenteritis.
FEATURES Location/Qualifiers
source 1. .302100
/organism="Vibrio parahaemolyticus"
/mol_type="genomic DNA"
/strain="O3:K6"
/sub_strain="RIMD 2210633"
/db_xref="taxon:670"
/chromosome="2"
misc_feature 53. .1139
/note="disrupted by frameshift, similar to part of hypothetical protein GB:AAF96452.1 (AE004386) percent identity 62 in 165 aa"
gene complement(1224. .2012)
/gene="VPA0861"
CDS complement(1224. .2012)
/gene="VPA0861"
/note="similar to GB:AAF96483.1 (AE004389) percent identity 73 in 261 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC62204.1"
/db_xref="GI:28809162"
/translation="MKVWIKASALLTLAGLTACSASPTGRNQILMFSDQEMSSLGAKS FDQMKKDIPISKDKKTNAYVQCVAKQITNAIPPQAGFKDWEVVVFDSDQVNAFALPGG KIGVYTGLLKVAKNQDQLATVIGHEVAHVLAEHSNERLSQSQLANAGLSLANVAIGAS EYKQYQQLTMAALGVGVQYGVILPYGRTQESEADIVGLEYMAKAGFNPNQSVDLWKNM SAASGGAQPPEFFSTHPSHSTRIKDLQATINKLPKYNVKAPKCG"
gene 2217. .2651
/gene="VPA0862"
CDS 2217. .2651
/gene="VPA0862"
/note="similar to GB:AAF96482.1 (AE004389) percent identity 75 in 144 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC62205.1"
/db_xref="GI:28809163"
/translation="MSIWKKPISLDILNATSRNTMMEHLQIEYTNFTDDSLSATMPVC SFTHQPLGMLHGGASVVLAETLGSIAANFCVSEGNYCVGLDINANHVRAMRSGYVIGT AKPIHLGVSTQVWQIEITDERDRLVCTSRLTIAVKQQKAKLA"
gene 2694. .2927
/gene="VPA0863"
CDS 2694. .2927
/gene="VPA0863"
/note="similar to GB:AAF96481.1 (AE004389) percent identity 56 in 76 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC62206.1"
/db_xref="GI:28809164"
/translation="MVIEFSLGKIIATQREIVIKLTGEAMVTLQAQTDALQLLGRGAN VVLAHGAECKWSIKLDNEEQLIQLAQEIGIDIQ"
gene 3007. .4713
/gene="VPA0864"
CDS 3007. .4713
/gene="VPA0864"
/note="similar to GB:AAF96480.1 (AE004388) percent identity 66 in 565 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative ABC transporter periplasmic solute binding protein"
/protein_id="BAC62207.1"
/db_xref="GI:28809165"
/translation="MSSPRLRVQFETLFEKFAGNDTEVQLEDITDALFCTRRNARIVL NKLEEEGWIEWHPAAGRGKLSKLIFKRNRSDVSENLARRYLDEGKIGQALDALDNDAA RLTQVIQGYLGLQHRQGEQVVRLPYYRPLAMLNPQKPMRRSEQHIARQVFSGLTRLDE SEQLQPDLAHTWEALSDTHWRFYLRRGVRFHNGEPLTTDCVVESILALSGLNLFSHIQ QVSSPQPWTVDIKLVRPDRYLPLALSESQAKVLLPQALRSEDFDRKPIGTGPFQVKVN DDKRLILTAFDGYFGFRPLLGQVEVWVIDEAYSSMVYPSLSKPKMNKQASTDEVELDP GCTFLLLNKNTGIAKDPRWAEFLSQTLNSHQIYAHVPQDKVIELGVLQAFGLKPGWID LRPTSVGEVPQQGKSITVAYQKKHPMFPVIAKAIKHLLKPHSVDVEFIRYDTQPPNPE EVDVWVKAMGIATNRNDALAGWLLDYSDIEKFSAGYDFSEWATLVDQWRAGIHADFPA RELGRQLVKSCQVIPMFHCWLGVNKDHSGALQNAKCNALGWFDFSNVWVKPEIEDNGQ AE"
gene complement(5251. .5847)
/gene="VPA0865"
CDS complement(5251. .5847)
/gene="VPA0865"
/note="similar to GB:AAF96687.1 (AE004407) percent identity 71 in 192 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC62208.1"
/db_xref="GI:28809166"
/translation="MKSTSTLLGLTIFYAIVFLFSALEPSSRAVWFAEIIPAIGILIA IWAISIRYQFSNTAYLLMFIWLCLHTIGAKYTFAEVPFDWFNNLIGSERNNFDRVAHF AIGLYAYPIAEYLIRNKKFNPTFSCWFALFAIMSLAAGYEIIEWWYAELAGGDEGIAF LGSQGDIWDAQKDMLCDTTGAILSLFLMSAQRRFAKPF"
gene 6123. .6962
/gene="VPA0866"
CDS 6123. .6962
/gene="VPA0866"
/note="similar to GB:AAK22187.1 (AE005694) percent identity 30 in 298 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative oxidoreductase iron/ascorbate family"
/protein_id="BAC62209.1"
/db_xref="GI:28809167"
/translation="MKLETVDYLADDAAQQFVASLRETGFGVLKNHPIPQELVESIYK NWHEFFCSEEKNDFQFNVETQDGYFPPSVSEVAKGHKVKDIKEYFHVYPWGQIPEQLK AEILEYYERANAFAQELLGWVEEYAPKDVQEKFSIALSEMINNSDKTLLRVLHYPPMT GEEEPGAIRAAAHEDINLLTVLPAANEPGLQVMSKDGEWIDVPCDFGNLIINIGDMLQ EASGGYFPSTTHRVINPTGARQEKSRISLPLFLHPKPDTVLSERYTAHSYLMERLREL GVI"
gene 7095. .7715
/gene="VPA0867"
CDS 7095. .7715
/gene="VPA0867"

```
                     /note="similar to GB:AAF96686.1 (AE004407) percent
                     identity 63 in 204 aa"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="DnaJ-related protein"
                     /protein_id="BAC62210.1"
                     /db_xref="GI:28809168"
                     /translation="MEDQYGVCDSSQSEIENPLLWPMLEILKRQPSGWKVHTLASKLS
                     ERGYINQLDPSPDKDLFKRNFLIMNALYQLQETLFPDSWLQVEAMDIMLMSALEASRN
                     SVDPNDPLRQYYLDWSHYEANEDEVRRLLNEFWTRYQRFVGGASEHNIDKAKALKLFD
                     LAGDVSRTEIRRQWRKLALRWHPDRENGDAERFRVLCEAWNVLRNG"
     gene            complement(7844. .9499)
                     /gene="VPA0868"
     CDS             complement(7844. .9499)
                     /gene="VPA0868"
                     /note="similar to GB:AAF96684.1 (AE004407) percent
                     identity 78 in 548 aa"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="BAC62211.1"
                     /db_xref="GI:28809169"
                     /translation="MDQLTATLKKIEKQNYRAYQQIKGQYDFTDFTLFIDHVQGDPYA
                     SASRFRATRAWSLTGLEWLKDESPAFQRAARDFIARSFEQFAKQENTVSIALNGQTVL
                     DSTAVLFTEEGIELRFRVNLPAEGRSVLGKKANNILTFHLPKFIRRATLERELDKEAM
                     VKHCQVVEDQSALREQLEAHNLVAFVANGSVLPRIAGNCDLPMKEAVEFTAPESLQVT
                     LHAPNKGYVTGLGIPKGITLIVGGGFHGKSTLLNAIERSIYDHIPGDGREYIVTDQKA
                     MKIRAEDGRCVHHLNLSNYINHLPMGKDTADFSTQDASGSTSQAAWLQESIEAGATSL
                     LIDEDTSATNFMIRDERMQALVAKGDEPITPLVDRIGQLRDELDISTIIVMGGSGDYL
                     DVANTVIQMHDYQAVDVTEKAKQVIAQHPTQRHNESEESLQTFRPRALNRVALMNILT
                     DGKFRVSAKGKDSLRFGKEFTDLSALEQIESADEVNAIGWLWFQLAQLPGWCNNPAKE
                     IEEMLSGEWHASLPKQGDLAKPRTLDVMAALNRMRKSQFKPSH"
     gene            complement(9571. .11481)
                     /gene="VPA0869"
     CDS             complement(9571. .11481)
                     /gene="VPA0869"
                     /note="similar to GB:AAF96683.1 (AE004406) percent
                     identity 50 in 612 aa"
                     /codon_start=1

  Query Match              3.0%;  Score 108.8;  DB 1;  Length 302100;
  Best Local Similarity   59.6%;  Pred. No. 1.9e-12;
  Matches  201;  Conservative    0;  Mismatches  133;  Indels    3;  Gaps    1;

Qy       3246 ACTCACAACATAAGCGCGTTAGTAAGAACCGTGTCAGCATCACCTATGACGTTGAAACGA 3305
             |||| ||| |   || ||  | |||   |||||   ||| ||||| || ||||||||
Db     192447 ACTCGAAACTTTCACGAGTACGGAAGCCTCGTGTTCACATTACCTACGATGTTGAAACCG 192388

Qy       3306 ATGGCGCCGTAAAGACGAAAGAGCTGCCGTTTGTTGTTGGCGTCATTGGCGACTTTTCAG 3365
             | ||  |      ||  |||||||| || ||||| || || || || |||||||||| |||
Db     192387 AAGGAACAACTTTGAAAAAAGAGCTTCCTTTTGTGGTGGGTGTTATGGGCGACTTTGCAG 192328

Qy       3366 GACACAAACCAGAATCAGAAAAAGTTGATTTAGAAGAGCGAGAGTTCACGGGTATCGATA 3425
             | || ||  | ||| ||   |||       ||| | || ||    ||||     ||||||
Db     192327 GCCAAAATACCGAAGCATTGAAA---CCATTAAAGGATCGCCGTTTCATTCAAATCGATC 192271

Qy       3426 AAGACAACTTCGATACAGTGATGGGGCAAATTCACCCGCGTCTTTCGTACAAGGTTGATA 3485
               || |||||||||   ||  |      | ||    ||      ||  || ||| || |  |
Db     192270 GTGATAACTTCGATGATGTACTTAAAAAGATGAGTCCATCGGTTAAGTTCAAAGTCGCAA 192211

Qy       3486 ACAAGCTTGCTAATGATGATAGCCAGTTTGAAGTGAACTTGAGCCTCCGTTCGATGAAAG 3545
             |||||  | || ||||| |||     | ||||  ||  |  ||||| |     |||||| |||
Db     192210 ACAAGTTAGCAAATGACGATTCTGAATTTGCGGTTGAACTGAGCTTTAAATCGATGCAAG 192151

Qy       3546 ATTTCCACCCAGAGAACTTAGTTGATNAAATTGAGCC 3582
             | ||  | ||||      | ||  || || | || ||
Db     192150 ACTTTGAACCAGCAGCGATTGTCAATCAAGTAGAACC 192114

RESULT 2
AP005348/c
```

```
LOCUS       AP005348              248850 bp    DNA     linear   BCT 04-DEC-2003
DEFINITION  Vibrio vulnificus YJ016 DNA, chromosome II, complete genome,
            section 5/8.
ACCESSION   AP005348 BA000038
VERSION     AP005348.1  GI:37201116
KEYWORDS    .
SOURCE      Vibrio vulnificus YJ016
  ORGANISM  Vibrio vulnificus YJ016
            Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
            Vibrionaceae; Vibrio.
REFERENCE   1
  AUTHORS   Chen,C.Y., Wu,K.M., Chang,Y.C., Chang,C.H., Tsai,H.C., Liao,T.L.,
            Liu,Y.M., Chen,H.J., Shen,A.B., Li,J.C., Su,T.L., Shao,C.P.,
            Lee,C.T., Hor,L.I. and Tsai,S.F.
  TITLE     Comparative Genome Analysis of Vibrio vulnificus, a Marine Pathogen
  JOURNAL   Genome Res. 13, 2577-2587 (2003)
REFERENCE   2  (bases 1 to 248850)
  AUTHORS   Chen,C.Y., Wu,K.M. and Tsai,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2002) Shih Feng Tsai, National Health Research
            Institutes, Division of Molecular and Genomic Medicine; 128,
            Yen-Chiu-Yuan Road, Sec 2, Taipei, Taiwan 115, Republic of China
            (E-mail:petsai@nhri.org.tw, Tel:886-2-8146-1041,
            Fax:886-2-2789-0484)
COMMENT     This sequence was determined by the Sequencing Core of the National
            Yang-Ming University Genome Research Center (YMGC;
            http://genome.ym.edu.tw).
FEATURES             Location/Qualifiers
     source          1. .248850
                     /organism="Vibrio vulnificus YJ016"
                     /mol_type="genomic DNA"
                     /strain="YJ016"
                     /db_xref="taxon:196600"
                     /chromosome="II"
     gene            383. .1786
                     /gene="VVA0914"
     CDS             383. .1786
                     /gene="VVA0914"
                     /codon_start=1
                     /transl_table=11
                     /product="anaerobic C4-dicarboxylate transporter"
                     /protein_id="BAC96940.1"
                     /db_xref="GI:37201117"
                     /translation="MNRVYYFFCITRENEITNISLRINNMLYLEFLFLLLMLYIGSRY
                     GGIGLGVVSGIGLVIEVFVFKMPPTSPPVTVMLIILAVVTCASILEAAGGLKYMLQVA
                     ERMLRKNPKRVTLIAPFVTYTMTFMLGTGHAVYSIMPIIGDVALKNGIRPERPMAAAS
                     VASQIAITASPISAAVVYYLAQLSDINHEITLLSILLVTVPATLFGTLLMSLYSMKRG
                     KELEDDPEYQARLKDPVWREKILNTTATSLDEHLPAAARNSVLLFILSILTIVVIAMI
                     PEVRIIGEGTKPISMAVVIQMMMLCFGGIILLATKTDPRSVPNGVVFKSGMVAAIAIF
                     GIAWMSDTYFQYAMPQFKSGIVEMVNNYPWTFALALFIVSVVVNSQAAVARMMLPVGL
                     GLGLEPALLVGLMPALYGYFFIPNYPSDIATVNFDNTGTTKIGKWYFNHSFMSVGLIG
                     VIGACCLGYLLGQMIIG"
     gene            1858. .1891
                     /gene="rpt2T08"
     repeat_region   1858. .1891
                     /note="2 units of 17 bp in-tandem repeat"
                     /rpt_type=tandem
                     /rpt_unit="TTTATTTGTCGCTAAAC"
     gene            complement(1905. .3185)
                     /gene="VVA0915"
     CDS             complement(1905. .3185)
                     /gene="VVA0915"
                     /note="identified by GeneMark and Glimmer2"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="BAC96941.1"
                     /db_xref="GI:37201118"
                     /translation="MMKKRILTLLLTQWLWQPAVLAASGTNVSQYTARVMQNAQQLAD
                     NNALADAIAKLNSAEVSRPDDVAALSKLLGIYYWQAEQFTRSIASLEKALSFELFTFE
                     EEWVTRRMLTQLYLTQGDYKKAIPQLNKLLKQVPENESAAEIWRYLAQAQYSVENWRE
                     TIVATKGFQRVEKKPDTTILSLALAAHVQLEQWKSVIPLAKQLIALEPSNKNWWLQAY
                     NAYLNLGQEKSALDMLTLIELQGLELNEAETKSLAYLYAAQGIYQKAAQTLARIENSQ
```

QDPKLIKLQAQYWQAAKEWQQALELWQRLAKQESQYQWEVALLENQLQRYQQVLISLA
EMDNPKRQFDAQLLKINALYRLERIDQAYAQAKKADALRSTKQTQNWLKFLSHKLKSS
ESKLDKAKGEQVSEREHLTQQRNVVADKSADSSS"

gene complement(3182. .3820)
/gene="VVA0916"

CDS complement(3182. .3820)
/gene="VVA0916"
/codon_start=1
/transl_table=11
/product="periplasmic protein TonB2"
/protein_id="BAC96942.1"
/db_xref="GI:37201119"
/translation="MMRLVIAFPVALALVFALFSAMALLVNSGKKQAIDQTPSLSFNM
VMVEQDTEVTRRQRSVPPPPEIPEPPQPETPKMTQTPMQVTSTQVATPNIDMRFNVSG
LAVSIPATVSAPTQNASNTTKQLMPLSRVEAVYPAKAKKRRIEGYVILKFDIDETGRT
KNIEVVEAQPARFFERNAMDAIKRWRYQPQIVDGQAQTIYGYTTKIEFKMAQ"

gene complement(3820. .4224)
/gene="VVA0917"

CDS complement(3820. .4224)
/gene="VVA0917"
/codon_start=1
/transl_table=11
/product="biopolymer transport protein ExbD2"
/protein_id="BAC96943.1"
/db_xref="GI:37201120"
/translation="MRLGRRQSKSDEAQIDLTSMLDIVFIMLIFFIVTSSFVRESGVE
VNRPQAANVSAQKDAGIFVAITAANDIYIDKRMVDVERVQASIERLLLEQPDASLVIQ
ADEYAYNGVVVKVMDAAKAAGVKNIALAAEKS"

gene complement(4221. .4787)
/gene="VVA0918"

CDS complement(4221. .4787)
/gene="VVA0918"
/codon_start=1
/transl_table=11
/product="biopolymer transport protein ExbB2"
/protein_id="BAC96944.1"
/db_xref="GI:37201121"
/translation="MIIMDLHSDIWPELLTSSSWWNALYSFMQQGGEVLWWLAAVAVI
FWLLTIEKLLFIYCAYPRLRAETLQKWQMRHDHLSWFARSQREAWLDDAHQQLFKHLA
FIKVLVAICPMLGLLGTVTGMISVFDVMATQGSSNPKLMASGISLATLPTMAGMVAAL
VGMFVHARLVKLCQNREVKLEQSLRSDS"

gene complement(4784. .6151)
/gene="VVA0919"

CDS complement(4784. .6151)
/gene="VVA0919"
/codon_start=1
/transl_table=11
/product="biopolymer transport protein ExbB-related protein"
/protein_id="BAC96945.1"
/db_xref="GI:37201122"
/translation="MNKCRLISLALLVSLSLPSFAQTQLVPKATQESRQQQAENQQRE
NAFLSQEQQLAELKRQLEKQKSTLEEKNALLSSQFSDNEVTLSHLEEELRVETGSLGE
LFGVVRQHAKELKADLNDAVTAIGQTEQVETVNQIVAAQSLPDIAQLKTLWHVMQQQI
VASGEMRALQVNLINAEGAVEPTDVMALGNFALLAPQGFVNWKSDKGFAIPYAKLPEG
APTLDTLRSQHAISMVPVLIDPTRGVLLEQYAHSPSLADRFAAGGVVGKVIFALLAVG
LVIALYRGILLLTTKSKIRLQLKQPNQPGNNPLGRVLAVYDKDKPRSVEALELRLLET
IVDEQTGLEKGLSMLKLLAALAPMLGLLGTVIGMIETFQVITQFGNGDPKVMASGISM
ALVTTVLGLVAAMPLLLAHNVLSGWAEEIRSILEKQGIGLVAQQAEQQLPSLTPTMDK
VGNAA"

gene complement(6152. .6919)
/gene="VVA0920"

CDS complement(6152. .6919)
/gene="VVA0920"
/note="identified by GeneMark and Glimmer2"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC96946.1"
/db_xref="GI:37201123"
/translation="MSGYKAGAALLLWAVSFGAMSNDLEQAQKIQNRTHTAASKSQSV
INASADKSQLLQAEIEILKEEVRNLELYQRHLTSMIANQEQEKSSLAKQIQDIQVTRQ
GIVPLMYDMLDTLKTTINQDMPIKWQQRQERIEKLEQMMVRADVADAEKFRRILEAYQ
IELDYGSKLSAYQAPITLADGKQIEVEQLHLGRVSLIARTLNANQFWYWDQNQRHWLS
LTTIPSTDVNRAFDIAAQKVAPAMLNMPVALATAEAK"

gene complement(7182. .7670)
/gene="VVA0921"

CDS complement(7182. .7670)
/gene="VVA0921"
/codon_start=1
/transl_table=11
/product="uncharacterized conserved protein"
/protein_id="BAC96947.1"
/db_xref="GI:37201124"
/translation="MRVMTSKVAIFVDVQNIYYTCKEKYRKHFDYNRFWQSATQDKQV
VVAHAYAISSKDPGQRQFHHILRGIGFDVKLKPFIQRWDGSAKGDWDVGLALDAFEHA
PMVDEVILLSGDGDFEILVERIQQKFGKKVTVYGVPGLTAQNLQAVVDKYIPIEDELL
LP"

gene 7796. .8227
/gene="VVA0922"

CDS 7796. .8227
/gene="VVA0922"
/note="identified by GeneMark and Glimmer2"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC96948.1"
/db_xref="GI:37201125"
/translation="MRYWSIVLILFSALAHSYPADSEKWQHRSIMFFAPSHDQYVQQF
LLETLMEECSLQERDIVTLVIKEDGSTVPVWAQDAFDLSVLNAFYDIPKGQHTGILVG
KDGREKLRWGEKTDWTLLKKVIDQMPLRQQEMQRNPSRCAI"

gene complement(8234. .9076)
/gene="VVA0924"

CDS complement(8234. .9076)
/gene="VVA0924"
/codon_start=1
/transl_table=11
/product="uroporphyrinogen-III methylase"
/protein_id="BAC96949.1"
/db_xref="GI:37201126"
/translation="MLIKSKHFPHREALSGTLYHQIYSIGSTMTLTSLSSLQNQGFVS
LVGAGPGDPDLLTVKGYRVIQQAEVVVYDRLVSPEILALANEKAEMIYVGKKLDFHCV
PQEEINQILVRKALEGKRVVRLKGGDSFIFGRGGEELEELAEHGVRFEVVPGITAAAG
ATAYAGIPLTHRDHAQSVQFITGHIQKDGKEIEWPSLAQANNTLVFYMGLKQCRYIAE
KLLENGLDASTSCAIIENGTRREQRVLTGSLTQLPELATQAISPALIVVGSVTSLHDK
LKWF"

gene 8255. .8395
/gene="VVA0923"

CDS 8255. .8395
/gene="VVA0923"
/note="identified by Glimmer2"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC96950.1"
/db_xref="GI:37201127"
/translation="MQRSHAADHDQRGTDCLSGQFRELCQAARQYALLATGAIFNNRA

Query Match 2.9%; Score 105; DB 1; Length 248850;
Best Local Similarity 56.6%; Pred. No. 1.2e-11;
Matches 214; Conservative 0; Mismatches 161; Indels 3; Gaps 1;

Qy 3209 ATGAATATCTCATTTTAGGACACCGTTATGGCATTGAACTCACAACATAAGCGCGTTAGT 3268
Db 95004 ATGAGTAATTAAAGGGAGAACGATATGACTAGTATCCACTCGAAATTATCGCGAGTACGG 94945

Qy 3269 AAGAACCGTGTCAGCATCACCTATGACGTTGAAACGAATGGCGCCGTAAAGACGAAAGAG 3328
Db 94944 AAGCCACGTGTTCATATTACCTATGATGTTGAAACAGAAGGCACATCACTGAAGAAAGAA 94885

Qy 3329 CTGCCGTTTGTTGTTGGCGTCATTGGCGACTTTTCAGGACACAAACCAGAATCAGAAAAA 3388
Db 94884 CTTCCGTTTGTAGTGGGTGTGATGGGGGATTTTGCTGGCCAAAATACTGAAGCCCTCAAA 94825

Qy 3389 GTTGATTTAGAAGAGCGAGAGTTCACGGGTATCGATAAAGACAACTTCGATACAGTGATG 3448

```
Db      94824 ---CCACTAAAAGACCGTCGTTTTATTCAAATTGATCGAGACAACTTTGATGATGTACTC 94768

Qy       3449 GGGCAAATTCACCCGCGTCTTTCGTACAAGGTTGATAACAAGCTTGCTAATGATGATAGC 3508

Db      94767 AAGAAAATGAGTCCATCAGTGGAGTTTAAAGTGAAAAACACGATGGTGGATGACGGAACG 94708

Qy       3509 CAGTTTGAAGTGAACTTGAGCCTCCGTTCGATGAAAGATTTCCACCCAGAGAACTTAGTT 3568

Db      94707 GAATTTGCGGTCAATCTGCAGTTCCAATCGATGAAGGATTTCGAACCCGCCTCGATTGTT 94648

Qy       3569 GATNAAATTGAGCCGCTT 3586

Db      94647 AGACAAGTTGATCCGCTT 94630
```

```
RESULT 3
AE016809/c
LOCUS       AE016809              300883 bp    DNA     linear   BCT 29-SEP-2003
DEFINITION  Vibrio vulnificus CMCP6 chromosome II section 2 of 6 of the
            complete sequence.
ACCESSION   AE016809 AE016796
VERSION     AE016809.1  GI:27358288
KEYWORDS    .
SOURCE      Vibrio vulnificus CMCP6
  ORGANISM  Vibrio vulnificus CMCP6
            Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
            Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 300883)
  AUTHORS   Kim,Y.R., Lee,S.E., Kim,C.M., Kim,S.Y., Shin,E.K., Shin,D.H.,
            Chung,S.S., Choy,H.E., Progulske-Fox,A., Hillman,J.D., Handfield,M.
            and Rhee,J.H.
  TITLE     Characterization and Pathogenic Significance of Vibrio vulnificus
            Antigens Preferentially Expressed in Septicemic Patients
  JOURNAL   Infect. Immun. 71 (10), 5461-5471 (2003)
   PUBMED   14500463
REFERENCE   2  (bases 1 to 300883)
  AUTHORS   Jeong,H., Moon,Y.H. and Kim,J.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-2002) Genotech Corp., 461-6, Jeonmin-dong,
            Yuseong-gu, Daejeon 305-811, South Korea
REFERENCE   3  (bases 1 to 300883)
  AUTHORS   Rhee,J.H., Kim,S.Y., Chung,S.S., Lee,S.E. and Choy,H.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-2002) Department of Microbiology, Genome Research
            Center for Enteropathogenic Bacteria, Chonnam National University
            Medical School, Hak-1-Dong, Dong-Gu, Kwang-Ju 501-746, South Korea
FEATURES             Location/Qualifiers
     source          1. .300883
                     /organism="Vibrio vulnificus CMCP6"
                     /mol_type="genomic DNA"
                     /strain="CMCP6"
                     /db_xref="taxon:216895"
                     /chromosome="II"
     gene            269. .577
                     /locus_tag="VV20274"
     CDS             269. .577
                     /locus_tag="VV20274"
                     /codon_start=1
                     /transl_table=11
                     /product="Unknown"
                     /protein_id="AAO07239.1"
                     /db_xref="GI:27358289"
                     /translation="MILENKPNTSRQFLISNNLKQCDFIYFHEKTRFQILKAQTNSTV
                     TISGKKDFLKIGGVVELTITQGRAAIGICQRALQEKDFTVDASLMRIADVKEAERCVR
                     "
     gene            565. .2553
                     /locus_tag="VV20275"
     CDS             565. .2553
                     /locus_tag="VV20275"
                     /note="COG2199"
                     /codon_start=1
                     /transl_table=11
                     /product="FOG: GGDEF domain"
                     /protein_id="AAO07240.1"
                     /db_xref="GI:27358290"
                     /translation="MCSLKSLSIKNKLLLPILLFGTLIFLIAQGFHFWASYEKDQQLQ
                     NHRISILAYGAAQNLQASVLFNDQKVMQEILDSFSIDPSILIVELRSLEGGVLARYQR
                     EMPVNVDNHVLSLPYPFETFITYSQRIDVEQNAIATLSITASTQTLEQMLTRLLESSV
                     VFLAILLLSGSTLYLIVQKVVIDPVYKLNESMQDLVEGRIKGGSLVSHSSDEMGELVN
                     AFNAMSQRLALRDSQVHHTLDRLEHEKAFATDVIEAVQTALLVVNEHGDILHHNAAAR
                     MLFQSDAQRIAQHKLSELMLCKEPSFVANALTSKNPLEERLVECRLPTGSKLLIKIST
                     RILSKRRATLFSIQDVTEIETALNRQRIAAGVFENSQDGLVVLDHNGLITMTNPAVTR
                     MLGYPQEILIGKSLLKAFDWQQIQTIMPTIQESLEHYGQWQGEVLEKHYNGNLIPMFA
                     KVSRISKKVGVQEVFDMVIILSDLSGVKEMERLEYLAHHDALTGLANRAKLHRTLEDE
                     LKLCNFTHSEFALLYMDLDGFKFINDTFGHDAGDHVLKVVSERLLNEVRSGDLVARLS
                     GDEFVLLIKRADNHSVTLLSERILESLRQPIEYLNQTMQVGCSIGVKLVGPQEKDMER
                     ILKSADTAMYKAKKSGKGQAILIGSEVGEACLSDVTSE"
     gene            complement(2838. .4703)
                     /locus_tag="VV20276"
     CDS             complement(2838. .4703)
                     /locus_tag="VV20276"
                     /note="similar to mostly Fe transport; COG1629"
                     /codon_start=1
                     /transl_table=11
                     /product="Outer membrane receptor protein"
                     /protein_id="AAO07241.1"
                     /db_xref="GI:27358291"
                     /translation="MEGKRVKILVDGASQPGVFDGGPYSFINSSAVSVDPDMLKSVEI
                     VKGAASSLHGSDAIGGVVAFETKDPADFLKNGKVFGGQAKLSYSSEDKSFSEHVAVAK
                     RFDNVEALVAYTRRDGKELQNFSKAPYDDYSVESQDYVKNDLLIKLQSQLSDDHRLEF
                     LGEVIYNQNDSDIASSSYKNFNSEDTTKQNRLGLKHIWLVDTTITDSITSRLTWTDKE
                     ENGLTNRFKEASAGVPPRVPPNGDNQQKKDYQYAENKIELETQFDKELETHYLVYGLS
                     YKTSRISNTNREFNSDPNTPDKIYVYTPDAKEMSFGLFFQDEISVLNDKLVLTPGVRY
                     DYFSTDPSDTVDESFIKFSDSAFTGRLGATYKVAQEGTLFTQISQGFRAPSFNELYYT
                     YDNPGHGYTNRPNPNLESEKSLSYELGYRHNTIASTSEIALYYSDYTDFIEQTSTRGG
                     GLTEYTYVNISEATIKGIEVSNTLDLNRIANLPLGMTTRFVAAYTVGEDGEGNPLNSV
                     NPWNAVGAVNYDAPNTLWGTSIKVNYTAKKSNSHINLDGNNGGTANQVELPSATVIDL
                     TAYFKPTQDVTVRAGIMNLTNEKYYLWNDIRGEKELSRDLTQAERNYNISVKYEF"
     gene            5180. .6094
                     /locus_tag="VV20277"
     CDS             5180. .6094
                     /locus_tag="VV20277"
                     /note="COG0583"
                     /codon_start=1
                     /transl_table=11
                     /product="Transcriptional regulator"
                     /protein_id="AAO07242.1"
                     /db_xref="GI:27358292"
                     /translation="MILITKGVEVQSPITLEALHILDAIDRRGSFASAANELDRAPSS
                     LSYQIQKLEQDLDIMIFDRSGHKANFTDAGRLILERGRTILAATEKLVNDATLLANGW
                     ELDISIAYDGIIGIHHFFSLVDALGNVSKTRVRLQEEILAGCWESLQSGRADLLVCPK
                     LEVLPHDVKADTIGKMSMTWVAAPNHYVHKRSGEFNDTAREKYRIIAIADTAREQPAI
                     TVNIIQKQPRLTVTNLQAKVDALVTGLGIGTLPTDIATPLIEAGKLKEIEGAEQQHLE
                     IILAWKRNQMGEAKSWCIQYLKKNWKLK"
     gene            complement(6143. .8821)
                     /locus_tag="VV20278"
     CDS             complement(6143. .8821)
                     /locus_tag="VV20278"
                     /note="other COG candidate exists; similar to
                     acetyltransferase; COG0454"
                     /codon_start=1
                     /transl_table=11
                     /product="Histone acetyltransferase HPA2"
                     /protein_id="AAO07243.1"
                     /db_xref="GI:27358293"
                     /translation="MHKLQKLFKPTSIAVIGASQRHLSAGNIVMKNLLQSGFDGTIIP
                     VTPRYKSVAGVLAYPSVTELPLVPDIAILCTRSERNVALFKQLAEIGTEFVIVLASDT
                     DYPHQDDQSVADACLRIARESNMRVLGPNSLGLILPWQKFNASFSPVTAKPGKIAFIS
                     QSAAVCTTILDWANDKEIGFSAFVSIGNGLDIDFDELLDYLSTDSKTDAILLYVDSIT
                     DARRFISAARAASRNRRILVLKGGRSILGRRALNKSGADTLDIVYDSAIRRSGMLRVN
                     NTHELFAAVETLTHSVPLRGERLAIITNGAGPALMAVDTLFERGGQLANLSSETIEML
                     SQHLPSSWSHNNPIDIVGDADKKRYVATINAILDSDCADAVLIMHSPSAVSDSVETAQ
                     AIVDAIQSHPRHKRFNILTNWSGEQTAKPARMILTKAGIPTYRTPESAVVAFMHLVEY
                     RRNQKQLMETPTTAEPLHVSDVNAAKAWVSEHLMDKEIAQLDTHQLGSLLKLFHFDVL
                     PTWIASDTSEAVHIAEQIGYPVAVKLRSPDIAHKSDVQGVMLNLSNSQEVASAADAIL
                     DRTKLAYPSANIHGLLVQGMAKLAGGEEIRVKVITDKVFGPVILLGQGGSEWSEARDA
```

```
                TAALPPLNMTLARYQIVRAIKEGHIRLQKLPEPMDVLGLAKPLVRISQMVVELPQIKE
                LDIHPVLANGEHFTILDADLTLENYLGHGQERLAIRPFPAEFVETMQLKDGEPILLRP
                ILPEDEPSHAEFINNVSKEDLYKRFFSDVGEFNHEALANLTQIDYDREMAFVAVSFSS
                GEAKIIGVARALINPENTEAEFAILIRSDLKGKGLGKVLMNKIIDYCRQKGTKKMTGM
                TMPTNRGMLTLAQKLGFSLDIQFEDGVADMVLPLNP"
gene            9046. .9621
                /locus_tag="VV20279"
CDS             9046. .9621
                /locus_tag="VV20279"
                /note="COG3266"
                /codon_start=1
                /transl_table=11
                /product="Uncharacterized protein conserved in bacteria"
                /protein_id="AAO07244.1"
                /db_xref="GI:27358294"
                /translation="MKKVAIITLSVLLSACASDNYVANVTTESHREEFKIEPIPEPLM
                AQEETVQEVSIPQQVVKMEPQTEEKKVVKLSPQSETTKAVKIVPPSKKQQDKMQRFGY
                TVQVVAVGSQAKVDYFASKLPKDGQPIWENYKIVNGTKWFTVLYGDFATSTEAKKAIA
                QLPQEFKELQPFVKSIDAIKNSEYPTLNKLN"
gene            9788. .10783
                /locus_tag="VV20280"
CDS             9788. .10783
                /locus_tag="VV20280"
                /note="similar to ATP-grasp enzymes; COG1181"
                /codon_start=1
                /transl_table=11
                /product="D-alanine-D-alanine ligase"
                /protein_id="AAO07245.1"
                /db_xref="GI:27358295"
                /translation="MMRNVLLLCGGGSSEHEISLLSSEYLQQQLGLIENVNVLKVEIK
                NEGWFDQKERLVYLDIHTKSVKSDEFNESIHIDFIVPCIHGFPGETGDIQSLFELAGI
                PYLGCGPEASSNSFNKITSKLWYDAIGIPNTPYLFLTENNEESHQLSREAFDKWGKLF
                VKAARQGSSVGCYSVTKIEQLSDAIDKAFGFSHQVLVEKAVKPRELEVSAYEMNGQLH
                ISKPGEIIAPDGAFYSYDEKYSAGSHSITEVEAKNLTEQQLATIQLCSEKVFRQMNLR
                HLSRIDFFLTSEGEIYLNEVNTFPGMTKISMFPKMLQHNGHKFHEFLADCIERSL"
gene            11080. .11454
                /locus_tag="VV20281"
CDS             11080. .11454
                /locus_tag="VV20281"
                /codon_start=1
                /transl_table=11
                /product="Conserved hypothetical protein"
                /protein_id="AAO07246.1"
                /db_xref="GI:27358296"
                /translation="MAISHYRGFNLQSAPNDTNIWQVKIKNRVLQGSLSAVKKSIDWW
                CDTASIIDPREFSSLDRKQEATGSPQAENFNGFAIKNDTGEPNAWYCFFNGKLIKGSK
                AAIQRHIEAYLIAKQKAEQQKK"
gene            11457. .11768
                /locus_tag="VV20282"
CDS             11457. .11768
                /locus_tag="VV20282"
                /note="eIF-1/SUI1; COG0023"
                /codon_start=1
                /transl_table=11
                /product="Translation initiation factor 1 related protein"
                /protein_id="AAO07247.1"
                /db_xref="GI:27358297"
                /translation="MTLVYSTETGRIKPEEPKASRPKGDGIVRIQRETKGRKGKGVCV
                VTGLDLDDAPLKLLAAELKKICGCGGSVKDGNIEIQGDARDKIKTHLEKKGMKVKFAG
                G"
gene            12360. .12782
                /locus_tag="VV20283"
CDS             12360. .12782
                /locus_tag="VV20283"
                /note="COG0499"
                /codon_start=1
                /transl_table=11
                /product="S-adenosylhomocysteine hydrolase"
                /protein_id="AAO07248.1"
                /db_xref="GI:27358298"
                /translation="MDFEAFNVLFWRVFPSIKAGAEYFHVQPATVRRWLSGSIPINPM
```

Query Match 2.9%; Score 105; DB 1; Length 300883;

Best Local Similarity 56.6%; Pred. No. 1.2e-11;
Matches 214; Conservative 0; Mismatches 161; Indels 3; Gaps 1;

```
Qy       3209 ATGAATATCTCATTTTAGGACACCGTTATGGCATTGAACTCACAACATAAGCGCGTTAGT 3268
             |||| ||  | |   || ||   | |      | ||||  ||     ||| || |
Db     171496 ATGAGTAATTAAAGGGAGAACGATATGACTAGTATCCACTCGAAATTATCGCGAGTACGG 171437

Qy       3269 AAGAACCGTGTCAGCATCACCTATGACGTTGAAACGAATGGCGCCGTAAAGACGAAAGAG 3328
             |||   |||||    || ||||||||| ||||||||  | ||| |   |  || ||||||
Db     171436 AAGCCACGTGTTCATATTACCTATGATGTTGAAACAGAAGGCACATCACTGAAGAAAGAA 171377

Qy       3329 CTGCCGTTTGTTGTTGGCGTCATTGGCGACTTTTCAGGACACAAACCAGAATCAGAAAAA 3388
             || ||||||||| || || || || || || || ||| | || || ||  | ||| |    |||
Db     171376 CTTCCGTTTGTAGTGGGTGTGATGGGGGATTTTGCTGGCCAAAATACTGAAGCCCTCAAA 171317

Qy       3389 GTTGATTTAGAAGAGCGAGAGTTCACGGGTATCGATAAAGACAACTTCGATACAGTGATG 3448
                 || |||| ||    || |    || |||  |||||||| |||   || |
Db     171316 ---CCACTAAAAGACCGTCGCTTTATTCAAATTGATCGAGACAACTTTGATGATGTACTC 171260

Qy       3449 GGGCAAATTCACCCGCGTCTTTCGTACAAGGTTGATAACAAGCTTGCTAATGATGATAGC 3508
               | ||||   ||      |  || || ||  | |||| | | |   |||| |  |
Db     171259 AAGAAAATGAGTCCATCAGTGGAGTTTAAAGTGAAAAACACGATGGTGGATGACGGAACG 171200

Qy       3509 CAGTTTGAAGTGAACTTGAGCCTCCGTTCGATGAAAGATTTCCACCCAGAGAACTTAGTT 3568
               | ||||  || ||  ||   |||  |||||||| |||||| | || |     | |||
Db     171199 GAATTTGCGGTCAATCTGCAGTTCCAATCGATGAAGGATTTCGAACCCGCCTCGATTGTT 171140

Qy       3569 GATNAAATTGAGCCGCTT 3586
                 || |||| ||||||
Db     171139 AGACAAGTTGATCCGCTT 171122
```

```
RESULT 4
AE016785/c
LOCUS       AE016785               300395 bp    DNA     linear   BCT 12-DEC-2002
DEFINITION  Pseudomonas putida KT2440 section 12 of 21 of the complete genome.
ACCESSION   AE016785 AE015451
VERSION     AE016785.1  GI:26557029
KEYWORDS    .
SOURCE      Pseudomonas putida KT2440
  ORGANISM  Pseudomonas putida KT2440
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 300395)
  AUTHORS   Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
            Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
            Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
            Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
            Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T.,
            Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
            Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
            Duesterhoft,A., Tummler,B. and Fraser,C.
  TITLE     Complete genome sequence and comparative analysis of the
            metabolically versatile Pseudomonas putida KT2440
  JOURNAL   Environ. Microbiol. 4 (12), 799-808 (2002)
REFERENCE   2  (bases 1 to 300395)
  AUTHORS   Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
            Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
            Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
            Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
            Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T.,
            Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
            Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
            Duesterhoft,A., Tummler,B. and Fraser,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .300395
                     /organism="Pseudomonas putida KT2440"
                     /mol_type="genomic DNA"
                     /strain="KT2440"
                     /db_xref="taxon:160488"
```

```
gene            88. .1707
                /gene="mqo-3"
                /locus_tag="PP2925"
CDS             88. .1707
                /gene="mqo-3"
                /locus_tag="PP2925"
                /note="similar to GB:L19401, GB:D16391, GB:Z22957,
                PID:307296, PID:312050, GB:L19401, GB:D16391, GB:Z22957,
                PID:307296, and PID:312050; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="malate:quinone oxidoreductase"
                /protein_id="AAN68533.1"
                /db_xref="GI:24984529"
                /translation="MKITIKPKLAMAGLALAIGMLNAQAAETRKVDVLLVGGGIMSST
                LAVWLNELEPGWSMEMVERLDKVAEESSNGWNNAGTGHSALAELNYTPEKDGKIDISK
                AVEINESFQVTRQFLAWQVKQGVLKNPHSFINTTPHMSFVWGDDNIRFLKKRYEALQA
                SPLFKPMQYSEDHEQIRKWVPLMMEGRDPNQKLAVTWTPIGTDVNFGEITRQYVGYLQ
                TRPNFDLKLSSEVQDITRNDDGSWHVEYKNLKNGSKAATDARFLFIGAGGAALPLLQK
                SGIDEAKNYGGFPVGGSFLVTDNPAIAQRHMAKAYGIAATGAPPMSVPHLDTRVLDGK
                RLILFGPFATFSTKFLKQGSLLDLLASTSVHNAWPMVRVGVREFDLVQYLIGQVLQSD
                DDRFEALRTYFPEAKKEDWRLWQAGQRVQIIKKDEALGGVLKLGTEVVTSRDGSIAGL
                LGASPGASTAAPIMLDVLNKVFKNKVASPEWQAKIKQIIPSYGIRLNDHPDQLEKEWA
                YTNEVLQLEPPVSPQRPESIRPADSQGVASR"
gene            complement(1717. .3045)
                /locus_tag="PP2926"
CDS             complement(1717. .3045)
                /locus_tag="PP2926"
                /note="identified by match to PFAM protein family HMM
                PF00984"
                /codon_start=1
                /transl_table=11
                /product="UDP-glucose dehydrogenase, putative"
                /protein_id="AAN68534.1"
                /db_xref="GI:24984530"
                /translation="MKVTVFGTGYVGLTQAVCLAQVGHSVLCMDVDADRVASLSEGHC
                PIFEPGLAPLLEKNLACGRLRFTTDAAAAANYARLQFIAVGTPPQADGSADLKHVFAV
                VDSILEHADGPKVIVNKSTVPVGTVHRIKARIAQAVADTSRFQVISNPEFLKEGSAVD
                DCMRPQRIIIGGAEAAEVELLRELYLPFNRNREKFMVMDARSAELTKYAANCMLATKI
                SFINEIANLAEHLGADIEMVRRGIGSDPRIGYDFIYAGCGFGGSCFPKDLQALRHSAE
                AEGFETQLLRAVESVNEQQKGRLFSKIQRHYPGGLRGKVFALWGLSFKPNTNDIREAS
                SRVLLEALWAAGARVQAHDPQAMEEIRRHYGPRADLQLVPCKDDALRGADALVIVTEW
                QDYRVLNLDEVPQQLADRVVFDGRNLYEPEHMASAGLAYYGIGRGKIAPA"
gene            3290. .4336
                /locus_tag="PP2927"
CDS             3290. .4336
                /locus_tag="PP2927"
                /note="identified by match to PFAM protein family HMM
                PF02390"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAN68535.1"
                /db_xref="GI:24984531"
                /translation="MDEARLHDFMGKLVGDMGAAATLANVILGDELGLYRAMADSQAV
                TPEQLAEKTGCHPRLVREWLNAQAASGYMVHDKGRFVLPEEQAMALALEDSPAYMAGG
                AAVVAALFHDKDKLVAAMRGDGGLAWGDHHPCMFSGTERFFRPGYRTFLVADWLPALE
                GVVAKLQAGAKVADVGCGHGASTLVMAQAFPASTFVGYDYHQPSIVTANARAREAGLE
                GRVSFQQASAKEYPGHDHDLVCFFDCLHDMGDPVGAARHAYHALKADGTVMLVEPYAE
                DSLEGNLTPVGRLFYAASTFICTPNSLSQEVGLGLGAQAGEARLRAVFEEAGFSRFRR
                ATQTPFNLILEARK"
gene            complement(4467. .5711)
                /locus_tag="PP2928"
CDS             complement(4467. .5711)
                /locus_tag="PP2928"
                /note="identified by match to PFAM protein family HMM
                PF02896"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAN68536.1"
                /db_xref="GI:24984532"
                /translation="MKKNVLIIGAGGVAKVVAHKCAQHNDELGRIAIASRNISKCQAI
                IDSVKAKGSLKVPADIQAFSLNALDVEATKALIRETESQIVINVGSAFLNMSVLRACI
                DTGVAYLDTAIHEEPGKICETPPWYGNYEWKHLEECQAKNITAILGVGFDPGVVNSYA
                KLAQQQYFDSIDSIDILDVNAGSHGKYFATNFDPEINFREFTGQVWSWQNSQWTSNTM
                FEVKRTDDLPVVGSQNLYLTGHDEVHSISKNLNVPNVRFWMSFGEHYINVFTVLKNLG
                LLSEQPVKTAEGLEVVPLKVVKAVLPDPASLAPGYTGKTCIGDLVKGTKDGQPREVFI
                YNVADHEEAYAETDSQGISYTAGVPPVAAALLVARGEWDAKRMVNVEELPAEPFLKAL
                DVMGLPTRVKDEKGDRPWDAEA"
gene            complement(5744. .6841)
                /gene="nspC"
                /locus_tag="PP2929"
CDS             complement(5744. .6841)
                /gene="nspC"
                /locus_tag="PP2929"
                /note="similar to GB:M80629, and PID:180492; identified by
                sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="carboxynorspermidine decarboxylase"
                /protein_id="AAN68537.1"
                /db_xref="GI:24984533"
                /translation="MIKTPYYLIDKTKLLSNMEKIAYVREHSGAKALLALKCFATWSV
                FDLMQQYMDGTTSSSLFELKLGRQKFAGETHAYSVAWADDEVEEMLENCDKIIFNSIG
                QLQRFAEQSEGKVRGLRVNPQVSSSDYLLADPARPFSRLGEWDPAKIEQVIEQISGFM
                FHNNCENGDFGLFDKMLTHIEERFGHLLHKVEWVSLGGGIHFTGEDYAVDAFCARLKA
                FSEKYGVQVYLEPGEAAITNSASLEVTVLDTLYNGKHLAVVDSSIEAHLLDLLIYRLN
                AKMAPNDGEHTYMVCGKSCLAGDIFGEYQFDRPLAIGDRLSFIDTAGYTMVKKNWFNG
                LKMPSIVVKQLDGSVEVVREFGFEDYVSSLS"
gene            complement(7027. .7944)
                /locus_tag="PP2930"
CDS             complement(7027. .7944)
                /locus_tag="PP2930"
                /note="similar to GP:57225, and SP:P17324; identified by
                sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="L-serine dehydratase, putative"
                /protein_id="AAN68538.1"
                /db_xref="GI:24984535"
                /translation="MTLHIHTPLIESRPLSLAAGRNIWLKLDALQPCGSFKLRGVGHA
                CEVHHARGARHFVSSSGGNAGLAVAYAGRKLGVPVTVVIPETTTERAKQLLHLEDAKV
                VVQGSSWQEANALAQTLVGPNDAFIHPFDDPLLWAGHASLVDEVAEAGLKPDAVVLSV
                GGGGLLSGVVEGLKRNGWGDVPVLAVETEGAASLHAAMQAGHSVELERIASVATSLGA
                KRVADQALMCVQQHPVHSHLVSDRAALEACERFLLDHRVLVEPACGAALAVAYDAKAL
                GALRNVLVVVCGGATATLEQIQAWLQQAD"
gene            complement(7941. .8765)
                /locus_tag="PP2931"
CDS             complement(7941. .8765)
                /locus_tag="PP2931"
                /note="similar to GP:5459226, and GP:5459226; identified
                by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein TIGR00148"
                /protein_id="AAN68539.1"
                /db_xref="GI:24984536"
                /translation="MLPRSRPPARAGWFRRARRRSAESYRARESRSRISTICVKMHKQ
                LILRSAMPIPPSEERQLTLTVLKAAIQALGSVAARNMEILLHDLDHPEHSVVAIVNGH
                LSGRSVGSPILAAPEQDQGFKALMQASADQHGREPVVLPDYPTTLKGRTLRSATAIFR
                DSSGHPFASLCVNTDVTGLDAAMNFLQQLQPLGAIPAVNEPADMELLMSEIIQAALQR
                SGQGRMNKQAKVEAVRVMQERGLFIVKGGVEKAASALGVTRYTIYNYLEQLRGASQ"
gene            complement(8698. .10401)
                /locus_tag="PP2932"
CDS             complement(8698. .10401)
                /locus_tag="PP2932"
                /note="identified by match to PFAM protein family HMM
                PF01425"
                /codon_start=1
                /transl_table=11
                /product="amidase family protein"
                /protein_id="AAN68540.1"
                /db_xref="GI:24984537"
                /translation="MIEVTEVSIAELRDALETGRTTAVELVKAYLARIDAYDGADTAT
```

```
                     ALNAVVVRNPEALKEAEASDARRAKGQVLSPLDGIPYTAKDSYLVKGLTAASGSPAFK
                     DLVAQRDAFTIERLRAAGAVCLGKTNMPPMANGGMQRGVYGRAESPYNADYLTAPFAS
                     GSSNGAGTATAASFSAFGLAEETWSSGRGPASNNGLCAYTPSRGVISVRGNWPLTPTM
                     DVVVPYARTMADLLEILDVVVADDADKRGDLWRLQPWVPIPAASTVRPASYLELAVDA
                     SALKGKRFGVPRMYINADPEAGTAEKPGIGGPTGQRINTRATVIDLWQQARQALEAAG
                     AEVLEVDFPLVSNCEGDRPGAPTVYNRGIVSKAFLHDELWELSGWAFDDFLRANDDPK
                     LNRLADVDGLKIFPHDPGTLPNREDDLAAGMDEYVNMAKRGLKTWDQIETLPDGLRGL
                     EQTRKLDLEDWMDKLGLDAVLFPTVADVGPADADINPVSADIAWSNGVWVANGNLAIR
                     HLGVPTVTVPMGVMADIGMPVGLTFAGKAYSDNALLSFAAAFEATGSRRMVPPRTPAL
                     G"
     gene            complement(10593. .11435)
                     /locus_tag="PP2933"
     CDS             complement(10593. .11435)
                     /locus_tag="PP2933"
                     /note="similar to GB:M59829, SP:P34931, SP:P48741, and

  Query Match             2.8%;  Score 99.4;  DB 1;  Length 300395;
  Best Local Similarity   58.1%;  Pred. No. 1.9e-10;
  Matches  194;  Conservative    0;  Mismatches  137;  Indels    3;  Gaps    1;

Qy       3252 AACATAAGCGCGTTAGTAAGAACCGTGTCAGCATCACCTATGACGTTGAAACGAATGGCG 3311
             |||   | |||||| |   |   || |||   |||||||| ||||| || |     | |
Db     171579 AACTCGACCGCGTTCGCGCGCCTCGGGTCCATATCACCTACGACGTGGACATCGGCGATG 171520

Qy       3312 CCGTAAAGACGAAAGAGCTGCCGTTTGTTGTTGGCGTCATTGGCGACTTTTCAGGACACA 3371
             || |  |||  || |||||||| || || || |||||  | || ||||| || ||  ||
Db     171519 CCATCGAGAAAAAGGAGCTGCCCTTCGTCGTCGGCGTACTGGGTGACTTCTCCGGCAACC 171460

Qy       3372 AACCAGAATCAGAAAAAGTTGATTTAGAAGAGCGAGAGTTCACGGGTATCGATAAAGACA 3431
               || ||  |           |  |  | || ||  |||||  |   ||||||   ||||
Db     171459 CGCCGGAGCC---GCTGCCCAAGCTCAAGGACCGCAAGTTCGTGTACATCGATCGCGACA 171403

Qy       3432 ACTTCGATACAGTGATGGGGCAAATTCACCCGCGTCTTTCGTACAAGGTTGATAACAAGC 3491
             |||||  |    ||| |   |   ||  | || || ||  | |||   || || ||||  |
Db     171402 ACTTCAACGGTGTGCTCAAGGGCATCAAGCCACGCCTGACCTACCGCGTCGACAACACCC 171343

Qy       3492 TTGCTAATGATGATAGCCAGTTTGAAGTGAACTTGAGCCTCCGTTCGATGAAAGATTTCC 3551
             | || ||  | | || ||||||  |  ||| |  ||| | ||       |  |||| |||
Db     171342 TGGCCAAGAACGGTACCCAGTTGGGCGTGGAGCTGAACTTCAACAGCCTCGAAGACTTCG 171283

Qy       3552 ACCCAGAGAACTTAGTTGATNAAATTGAGCCGCT 3585
              | ||| ||||| | ||  |  |  | ||||||||
Db     171282 AGCCACAGAACGTGGTCAAGCAGGTCGAGCCGCT 171249


RESULT 5
AE016924/c
LOCUS       AE016924               301457 bp    DNA     linear   BCT 14-OCT-2003
DEFINITION  Chromobacterium violaceum ATCC 12472 section 15 of 16 of the
            complete genome.
ACCESSION   AE016924 AE016825
VERSION     AE016924.1  GI:34105231
KEYWORDS    .
SOURCE      Chromobacterium violaceum ATCC 12472
  ORGANISM  Chromobacterium violaceum ATCC 12472
            Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
            Neisseriaceae; Chromobacterium.
REFERENCE   1  (bases 1 to 301457)
  AUTHORS   .
  CONSRTM   Brazilian National Genome Project Consortium
  TITLE     The complete genome sequence of Chromobacterium violaceum reveals
            remarkable and exploitable bacterial adaptability
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (20), 11660-11665 (2003)
  MEDLINE   22882880
   PUBMED   14500782
REFERENCE   2  (bases 1 to 301457)
  AUTHORS   Vasconcelos,A.T.R., de Almeida,D.F., Almeida,F.C., de
            Almeida,L.G.P., de Almeida,R., Goncalves,J.A.A., Andrade,E.M.,
            Antonio,R.V., Araripe,J., de Araujo,M.F.F., Filho,S.A., Azevedo,V.,
            Batista,A.J., Bataus,L.A.M., Batista,J.S., Belo,A., vander Berg,C.,
            Blamey,J., Bogo,M., Bonato,S., Bordignon,J., Brito,C.A.,
            Brocchi,M., Burity,H.A., Camargo,A.A., Cardoso,D.D.P.,
            Carneiro,N.P., Carraro,D.M., Carvalho,C.M.B., Cascardo,J.C.M.,
            Cavada,B.S., Chueire,L.M.O., Pasa,T.B.C., Duran,N., Fagundes,N.,
            Falcao,C.L., Fantinatti,F., Farias,I.P., Felipe,M.S.S.,
            Ferrari,L.P., Ferro,J.A., Ferro,M.I.T., Franco,G.R.,
            Freitas,N.S.A., Furlan,L.R., Gazzinelli,R.T., Gomes,E.A.,
            Goncalves,P.R., Grangeiro,T.B., Grattapaglia,D., Grisard,E.C.,
            Guimaraes,C.T., Hanna,E.S., Hungria,M., Jardim,S.N., Laurino,J.,
            Leoi,L.C.T., Fassarella,L., Lima,A., Loureiro,M.F., Lyra,M.C.P.,
            Macedo,M., Madeira,H.M.F., Manfio,G.P., Maranhao,A.Q.,
            Martins,W.S., di Mauro,S.M.Z., de Medeiros,S.R.B., Meissner,R.V.,
            Menck,C.F.M., Moreira,M.A.M., Nascimento,F.F., Nicolas,M.F.,
            Oliveira,J.G., Oliveira,S.C., Paixao,R.F.C., Parente,J.A.,
            Pedrosa,F.O., Pena,S.J.D., Perreira,J.O., Perreira,M.,
            Pinto,L.S.R.C., Pinto,L.S., Porto,J.I.R., Potrich,D.P.,
            Neto,C.E.R., Reis,A.M.M., Rigo,L.U., Rondinelli,E., dos
            Santos,E.B.P., Santos,F.R., Schneider,M.P.C., Seuanez,H.N.,
            Silva,A.M.R., da Silva,A.L.C., Silva,D.W., Silva,R., Simoes,I.C.,
            Simon,D., Soares,C.M.A., Soares,R.B.A., Souza,E.M., Souza,K.R.L.,
            Souza,R.C., Steffens,M.B.R., Steindel,M., Teixeira,S.R.,
            Urmenyi,T., Vettore,A., Wassem,R., Zaha,A. and Simpson,A.J.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2003) Labinfo, LNCC - Laboratorio Nacional de
            Computacao Cientifica, Rua Getulio Vargas 333, Petropolis, RJ
            25651070, Brazil
FEATURES             Location/Qualifiers
     source          1. .301457
                     /organism="Chromobacterium violaceum ATCC 12472"
                     /mol_type="genomic DNA"
                     /strain="ATCC 12472"
                     /db_xref="taxon:243365"
     gene            121. .1386
                     /gene="goaG"
                     /locus_tag="CV3926"
     CDS             121. .1386
                     /gene="goaG"
                     /locus_tag="CV3926"
                     /EC_number="2.6.1.19"
                     /note="identified by sequence similarity; putative; ORF
                     located using Glimmer/GeneMark/Blastx/COG0160"
                     /codon_start=1
                     /transl_table=11
                     /product="4-aminobutyrate transaminase"
                     /protein_id="AAQ61587.1"
                     /db_xref="GI:34105232"
                     /translation="MSQQQLMQRKSDATPRGVGVMCSFFADRAKNAEVWDTEGRRYID
                     FAGGIGVLNTGHLHPKVEAAVAEQLSRFSHTCYQVVPYESYISLAEKLNALAPGKHAK
                     KTAFFTTGAEAVENAVKIARAATGRPGVIAFGGAFHGRTLMGMALTGKVAPYKLGFGP
                     FPSDVYHAPYPNTLHGVSVQDSLVGLEKLFKFDIDPARVAAIIYEPVQGEGGFYAAPA
                     EWVKALRELCDRHGILLIADEVQSGFARTGKLLASEHFDAVPDLITMAKSLAGGFPLS
                     AVTGRAEVMDAPAPGGLGGTYAGNPLALAAAHAVLDVIDSEKLAERAQRLGDALKEKL
                     NGLKSAVPQIAEVRGLGAMIAVEFNQPGSHQPDPDFAKKVQQHALANGLILLTCGIYA
                     NVVRFLFPLTIEDEIFAEALGKLEAALKA"
     gene            1456. .2916
                     /gene="gabD"
                     /locus_tag="CV3927"
     CDS             1456. .2916
                     /gene="gabD"
                     /locus_tag="CV3927"
                     /EC_number="1.2.1.16"
                     /note="identified by sequence similarity; putative; ORF
                     located using Glimmer/GeneMark/Blastx/COG1012"
                     /codon_start=1
                     /transl_table=11
                     /product="succinate-semialdehyde dehydrogenase [NAD(P)]"
                     /protein_id="AAQ61588.1"
                     /db_xref="GI:34105233"
                     /translation="MLNLQDPSLLRQQCYLGGEWLDADGGQHIAVHNPATGEKIGQVP
                     KMGAAEAERAVAAAARAFPAWKAKSAKERAAILRRWFDLIIAHQQDLAIILTSEQGKP
                     LAEAKGEIAYGASYIEWYAEEAKRVYGDIVPGPAGDRRVLVTKEPIGVTAAITPWNFP
                     SAMITRKAAPALAAGCTMVVRPASQTPLSALALAELAERAGIPAGVFSVITGGSSEIG
                     AVLTGSDTVKKFSFTGSTEVGRKLIAQCAGTVKKVSMELGGNAPFIVFDDADLDAAVE
                     GALISKYRNAGQTCVCANRIYVQDGVYDAFAEKFAAAVATLKVGNGLEAGVSQGPLID
                     HNAVAKVEEHIADALAKGGRLIAGGKRHALGHTFFEPTLIADVTRDMKVAREETFGPL
                     APLFRFKTEEEAIRLANDTEFGLASYFYARDVGRIFRVAEGLEYGMVAVNSGLLSNEA
```

```
                APFGGVKQSGLGREGSKYGIEDYLEIKYVLLGGLDR"
gene            3010. .3330
                /locus_tag="CV3928"
CDS             3010. .3330
                /locus_tag="CV3928"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAQ61589.1"
                /db_xref="GI:34105234"
                /translation="MKYNHIGIPTSGRFDNEIDLPHLKMTVSDHQDNAFGIQWQRYWQ
                DAPYPELVKRVPHIAFEVEDLAQALEGHKLLIAPNSPSPGLTVAFIEVNGAPVELMQY
                DKNN"
gene            complement(3435. .3674)
                /gene="rRNA5S"
                /locus_tag="CVrRNA-5s-7"
rRNA            complement(3435. .3674)
                /gene="rRNA5S"
                /locus_tag="CVrRNA-5s-7"
                /product="rRNA-5s"
                /note="identified by sequence similarity; putative"
gene            complement(3675. .6614)
                /gene="rRNA23S"
                /locus_tag="CVrRNA-23s-7"
rRNA            complement(3675. .6614)
                /gene="rRNA23S"
                /locus_tag="CVrRNA-23s-7"
                /product="rRNA-23s"
                /note="identified by sequence similarity; putative"
gene            complement(6813. .6888)
                /locus_tag="CVtRNA-Ala-TGC-7"
tRNA            complement(6813. .6888)
                /locus_tag="CVtRNA-Ala-TGC-7"
                /product="tRNA-Ala"
gene            complement(6975. .7051)
                /locus_tag="CVtRNA-Ile-GAT-7"
tRNA            complement(6975. .7051)
                /locus_tag="CVtRNA-Ile-GAT-7"
                /product="tRNA-Ile"
gene            complement(7295. .8768)
                /gene="rRNA16S"
                /locus_tag="CVrRNA-16s-7"
rRNA            complement(7295. .8768)
                /gene="rRNA16S"
                /locus_tag="CVrRNA-16s-7"
                /product="rRNA-16s"
                /note="identified by sequence similarity; putative"
gene            8828. .9085
                /locus_tag="CV3929"
CDS             8828. .9085
                /locus_tag="CV3929"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAQ61591.1"
                /db_xref="GI:34105236"
                /translation="MQCKYSAFAKHLHLSVIHSVKEQCRSLRRRCGTVSFPLFRSLRQ
                LRRRTIRPDPKPVNTFREKNLRKMPILFQNRNNKLKRKHLC"
gene            9220. .10809
                /gene="nadB"
                /locus_tag="CV3930"
CDS             9220. .10809
                /gene="nadB"
                /locus_tag="CV3930"
                /EC_number="1.4.3.16"
                /note="identified by sequence similarity; putative; ORF
                located using Glimmer/GeneMark/Blastx/COG0029"
                /codon_start=1
                /transl_table=11
                /product="L-aspartate oxidase"
                /protein_id="AAQ61592.1"
                /db_xref="GI:34105237"
                /translation="MLKFDVLIIGSGLAGMTLALHLAETRKVGLITKRDLLEGNSTYA
                QGGIAAVLDTEDSVEDHIRDTLVAGAGLCNEKATRFIVEHSKEAIDWLIKLGVPFTRD
                DEHQTGYHLTREGGHSHRRIIHAADATGAAVTSTLGQKIKAHPNITELEHYIAIDLIT
                SQKLGQSGKRAYGVYALDKNGDEVVTIAAQHTILASGGAGKVYLYTTNPDTATGDGIA
                MGWRAGCRVGNMEFIQFHPTCLYHPHAKSFLISEAVRGEGGILKLPDGTRFMPQHDPR
                GELAPRDVVARAIDFEMKKHGLDCVYLDISYKPAQFIKEHFPNIYAHCLELGIDITQQ
                PIPVVPAVHFTCGGIVTDLAGRTDVDGLYAVGEVACTGLHGANRLASNSLLECLVIGK
                AAAEDILAADSIKLPAIPAWDDSRVTDPDEEVVISHNWDELRRAMWDYVGIVRTNKRL
                ERALHRIDLLKGEIDEYYSHFHVSNDLIELRNLVQTAELIVRSALLRKESRGLHYSRD
                YPNLLPEAAETVLAPDAPLGG"
gene            11124. .12206
                /locus_tag="CV3931"
CDS             11124. .12206
                /locus_tag="CV3931"
                /EC_number="3.2.1.132"
                /note="identified by sequence similarity; putative; ORF
                located using GeneMark/Blastx"
                /codon_start=1
                /transl_table=11
                /product="probable chitosanase A"
                /protein_id="AAQ61593.1"
                /db_xref="GI:34105238"
                /translation="MKYAAKRLSFPSAILAGLALLAAAGSAAAQGSTAGSPERQTVRL
                PDCAAAWSPASPYQAGSVASRGGVNYTAAFWTQGNPPEQGQAWQAGKRCRPAAQAKAA
                DHDANFPPATLKFLKANTGLDGEQWDNIMKLVNKPEQDSLEWTKFYGYCENIGDKRGY
                TIGIFGATTGGPNDEGPDGPTLFKEFDASSGASNPSIAGGLARAGVHGSMQGKILKIS
                DSAKVFCDKIGSLQNNPAWRDAMWNTFYKVYIQYSVQQARQRGFSSALTIGSFVDTAL
                NQGATGDSGTLQGLLSRSGNSGDEKTPMTAFYAQRSKIVDTNDYNQPPNGKNRVKQWS
                TLLNMGETDLKNADAAVQKVTNWEMK"

  Query Match             2.6%;  Score 91.6;  DB 1;  Length 301457;
  Best Local Similarity   56.3%;  Pred. No. 8.4e-09;
  Matches  191;  Conservative    0;  Mismatches  145;  Indels    3;  Gaps    1;

Qy       3250 ACAACATAAGCGCGTTAGTAAGAACCGTGTCAGCATCACCTATGACGTTGAAACGAATGG 3309
             |  |||    | | ||  ||  |   || ||    ||||||||||||||| || |     |
Db      62477 AAAACTGTCGAGGGTCCGTCCGCCGCGCGTGCAGATCACCTATGACGTCGAGATCGGCGA 62418

Qy       3310 CGCCGTAAAGACGAAAGAGCTGCCGTTTGTTGTTGGCGTCATTGGCGACTTTTCAGGACA 3369
             |||  |  | || || |||||||||| ||  | |||||  | ||||| | ||| || ||
Db      62417 CGCGATCGAAACCAAGGAGCTGCCGTTCGTGCTGGGCGTGCTGGGCGATTATTCCGGCCA 62358

Qy       3370 CAAACCAGAATCAGAAAAAGTTGATTTAGAAGAGCGAGAGTTCACGGGTATCGATAAAGA 3429
             ||    ||  |          | |  | |||||  |||||      |||||    ||
Db      62357 CAGCAAGGAGCCGCTGCCGA---AGATGAAGGAGCGCAAGTTCGTCCAGATCGACCGCGA 62301

Qy       3430 CAACTTCGATACAGTGATGGGGCAAATTCACCCGCGTCTTTCGTACAAGGTTGATAACAA 3489
             |||||||||    ||| ||  |   ||    ||||| ||  ||  |||||| ||| |||
Db      62300 CAACTTCGACGAGGTGCTGAAGGGCATGGCGCCCGCGCCTGGCGATGAAGGTGGACAACGC 62241

Qy       3490 GCTTGCTAATGATGATAGCCAGTTTGAAGTGAACTTGAGCCTCCGTTCGATGAAAGATTT 3549
             |||      | || |  | |||| | |  ||| | ||||    ||        ||   |||||
Db      62240 GCTGAAGGACGACGGCACCCAGCTGGGCGTGGAATTGAATTTCGAGAGCCTGTCGGATTT 62181

Qy       3550 CCACCCAGAGAACTTAGTTGATNAAATTGAGCCGCTTAA 3588
              | | ||  ||||| | |||.    | || || ||| | ||
Db      62180 CGAGCCGCAGAACGTGGTTCGCCAGATCGACCCGTTGAA 62142


RESULT 6
AL646080
LOCUS       AL646080              190050 bp    DNA     linear   BCT 11-APR-2003
DEFINITION  Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
            segment 5/11.
ACCESSION   AL646080 AL646053
VERSION     AL646080.1  GI:17431113
KEYWORDS    .
SOURCE      Ralstonia solanacearum
  ORGANISM  Ralstonia solanacearum
            Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
            Burkholderiaceae; Ralstonia.
REFERENCE   1
  AUTHORS   Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
            Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
```

```
            Chandler,M., Choisne,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
            Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
            Siguier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
            Weissenbach,J. and Boucher,C.A.
  TITLE     Genome sequence of the plant pathogen Ralstonia solanacearum
  JOURNAL   Nature 415 (6871), 497-502 (2002)
  MEDLINE   21681879
   PUBMED   11823852
REFERENCE   2  (bases 1 to 190050)
  AUTHORS   Boucher,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
            Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
            Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
            BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
            Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
            118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
            URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
            Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
            F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
            INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT     Christian.Boucher@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES             Location/Qualifiers
     source          1. .190050
                     /organism="Ralstonia solanacearum"
                     /mol_type="genomic DNA"
                     /strain="GMI1000"
                     /db_xref="taxon:305"
                     /plasmid="megaplasmid"
     gene            51. .362
                     /gene="RSp0643"
                     /note="synonym: RS05579"
     CDS             51. .362
                     /gene="RSp0643"
                     /function="miscellaneous; unknown"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="HYPOTHETICAL PROTEIN"
                     /protein_id="CAD17794.1"
                     /db_xref="GI:17431114"
                     /db_xref="GOA:Q8XS38"
                     /db_xref="SPTREMBL:Q8XS38"
                     /translation="MKTSPSPNPAAAGPASDPEAAGDGQVAFEQLHEYRNHTITIRPR
                     RIGGGLWRAAFAVSLDSRALVAVGTEVGQAFRTPIGAARDAVDAARSFIDARLKACRP
                     A"
     gene            397. .555
                     /gene="RSp0644"
                     /note="synonym: RS05578"
     CDS             397. .555
                     /gene="RSp0644"
                     /function="miscellaneous; unknown"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PUTATIVE SIGNAL PEPTIDE PROTEIN"
                     /protein_id="CAD17795.1"
                     /db_xref="GI:17431115"
                     /db_xref="GOA:Q8XS37"
                     /db_xref="SPTREMBL:Q8XS37"
                     /translation="MERQQLFAIAIVIAAVSAAYSLARDLNDDERDFPPGQAALVHGH
                     ATASAAGR"
     gene            complement(575. .1135)
                     /gene="RSp0645"
                     /note="synonym: RS05577"
     CDS             complement(575. .1135)
                     /gene="RSp0645"
                     /function="miscellaneous; not classified regulator"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE TRANSCRIPTION REGULATOR PROTEIN"
                     /protein_id="CAD17796.1"
                     /db_xref="GI:17431116"
                     /db_xref="GOA:Q8XS36"
                     /db_xref="SPTREMBL:Q8XS36"
                     /translation="MKRISTFPDFVDMKKVAIKFRSPAAAPPQGAALSALDRTDRAIL
                     KWLQRDASISNVALADKVSLSPPACLRRVERLKELGCIRGVVALLEPKALDLGTLVMI
                     GVVLDRSTPESFADFEKAVQKVPGCLECHVVTGEFDYFMLVRTRDNDSYNRLHAEQLL
                     YLPGVRQIRTFMTLKQVLSTTQLPIA"
     gene            1244. .2260
                     /gene="RSp0646"
                     /note="synonym: RS05576"
     CDS             1244. .2260
                     /gene="RSp0646"
                     /EC_number="4.1.99.4"
                     /function="miscellaneous; hypothetical/global homology"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE 1-AMINOCYCLOPROPANE-1-CARBOXYLATE
                     DEAMINASE (ACC DEAMINASE) PROTEIN"
                     /protein_id="CAD17797.1"
                     /db_xref="GI:17431117"
                     /db_xref="GOA:Q8XS35"
                     /db_xref="SPTREMBL:Q8XS35"
                     /translation="MNLNKHPRHPLTFGPTPIQPLKRLSAHLGGKVELYAKREDCNSG
                     LAFGGNKTRKLEYLVPEVLAGGYDTLVSIGGIQSNQTRQVAAVAAHLGLKCVLVQENW
                     VNYADAVYDRVGNIELSRILGADVRLDAAGFDIGIRPSWEQAMEDVRRAGGKPFPIPA
                     GCSEHPLGGLGFVGFAEEVRQQEAEFGFRFDYIVVCSVTGSTQAGMVVGFAADGRADR
                     VIGIDASAKPEQTREQILRIARDTAKLVELGRDITEDDVVLDTRYGGPEYGLPNEGTL
                     EAIRLCARQEGMLTDPVYEGKSMHGMIDRVRGGEFPEGSRVLYAHLGGVPALNAYSFL
                     FRNG"
     gene            complement(2336. .3508)
                     /gene="RSp0647"
                     /note="synonym: RS05575"
     CDS             complement(2336. .3508)
                     /gene="RSp0647"
                     /EC_number="4.2.1.17"
                     /function="miscellaneous; hypothetical/global homology"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PUTATIVE ENOYL-COENZYME A HYDRATASE PROTEIN"
                     /protein_id="CAD17798.1"
                     /db_xref="GI:17431118"
                     /db_xref="GOA:Q8XS34"
                     /db_xref="SPTREMBL:Q8XS34"
                     /translation="MSALPKPAEPAAAAEPEVLFHVVNRVAILTLNRPSALNALSHGM
                     VRQLSALLAQCRDDPRIAAVVLRGAGEKGFCAGGDVRALYRLARQGQAHGPDGWLQFF
                     VDEYRLDYALHRFPKPVVALLDGIAMGGGMGLGQGARLRVVTERSRIAMPETRIGFLP
                     DVGATHFLGAMPPEMALYIGLTGTTLSGADALHCGLADVCVPAEWLASFEDRLLRLSH
```

```
                     AGDPGDLMQALRRVFEPPCNIVPHAAIARTTPWIVRYFDRRSTIDRIVATLRESLERD
                     PAREARQWLQATLDAMATHSPTMLHVTREALLRGRQLTLAECFRMELGIVARAIEEGD
                     FCEGVRAHLVDKDHRPGWAPATLAGVRPERVRHFLSSPWRGAAHPLAHLGAQPQAA"
     gene            complement(3505. .4305)
                     /gene="RSp0648"
                     /note="synonym: RS05574"
     CDS             complement(3505. .4305)
                     /gene="RSp0648"
                     /EC_number="4.2.1.17"
                     /function="small molecule metabolism; degradation; carbon
                     compounds"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE ENOYL-COA HYDRATASE PROTEIN"
                     /protein_id="CAD17799.1"
                     /db_xref="GI:17431119"
                     /db_xref="GOA:Q8XS33"
                     /db_xref="SPTREMBL:Q8XS33"
                     /translation="MIALDWLHDGTVALLTLSRPPANAFTPEGLRQLQATVETLDADP
                     RVRALVITGDGPKFFSAGADLNTFASGDRDVAREAAARFGAAFEALQNARPVVIAAIN
                     GYAMGGGLECALACDIRIAERHAQLAVPETAVGLLCCGCGTQTLPWLVGEGWAKRMIL
                     TGERIDAQTALRIGLVEEVVDTGAAREAALAMARRVTTLSPRAVTFSKALIHQARHGV
                     PRTAALAVERERFVDLFGHPDQREGVNAFLEKRPPRWHTAARPEEALP"
     gene            complement(4302. .5198)
                     /gene="mmsB"
                     /note="synonyms: RS05573, RSp0649"
     CDS             complement(4302. .5198)
                     /gene="mmsB"
                     /EC_number="1.1.1.31"
                     /function="small molecule metabolism; amino acid
                     biosynthesis; isoleucine/valine"
                     /note="Product confidence : probable
                     Gene name confidence : putative
                     predicted by Codon_usage
                     predicted by Homology
```

Query Match 2.3%; Score 82.8; DB 1; Length 190050;
Best Local Similarity 56.3%; Pred. No. 6.1e-07;
Matches 175; Conservative 0; Mismatches 133; Indels 3; Gaps 1;

```
Qy       3275 CGTGTCAGCATCACCTATGACGTTGAAACGAATGGCGCCGTAAAGACGAAAGAGCTGCCG 3334
             || ||     | || || |||||||||| |   | |||  |  ||  ||| || ||||||
Db     112111 CGCGTGCAGCTGACGTACGACGTTGAAAAGGGCGACGCGATCGAGCAGAAGGAACTGCCG 112170

Qy       3335 TTTGTTGTTGGCGTCATTGGCGACTTTTCAGGACACAAACCAGAATCAGAAAAAGTTGAT 3394
             || || || |||||  | |  ||| | || || ||     | |||   |   |     |
Db     112171 TTCGTGGTCGGCGTGGTGGCGGACCTGTCGGGCCAGT---CGGAAGTGGCGCAGCCCAAG 112227

Qy       3395 TTAGAAGAGCGAGAGTTCACGGGTATCGATAAAGACAACTTCGATACAGTGATGGGGCAA 3454
               |    || ||  |||||      |||||    ||||||||||    ||||||  |
Db     112228 CTGCGCGACCGCAAGTTCGTCAACATCGACCGCGACAACTTCGACGACGTGATGTCGGCG 112287

Qy       3455 ATTCACCCGCGTCTTTCGTACAAGGTTGATAACAAGCTTGCTAATGATGATAGCCAGTTT 3514
                | |||||   | | |   ||||  | |||| |||  |  | || |   |||  ||
Db     112288 GCCGAGCCGCGCGCTGCCTTTCAGGTGCAGAACACGCTGACGGAAGAGGGCGGCCGCTTC 112347

Qy       3515 GAAGTGAACTTGAGCCTCCGTTCGATGAAAGATTTCCACCCAGAGAACTTAGTTGATNAA 3574
             |  ||  || | |   |||| ||| || | ||||||   || ||   | | ||  |  |
Db     112348 GGCGTCGACCTCAAGTTCCGCTCGCTGGAGGATTTCTCGCCGGAAGCCGTGGTGAACCAG 112407

Qy       3575 ATTGAGCCGCT 3585
             || |||||||||
Db     112408 ATCGAGCCGCT 112418
```

RESULT 7
AF361470

```
LOCUS       AF361470              33154 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Rhizobium leguminosarum bv. trifolii imp gene locus, complete
            sequence.
ACCESSION   AF361470
VERSION     AF361470.1  GI:16326458
KEYWORDS    .
SOURCE      Rhizobium leguminosarum bv. trifolii
  ORGANISM  Rhizobium leguminosarum bv. trifolii
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
REFERENCE   1  (bases 11567 to 12038)
  AUTHORS   Roest,H.P., Mulders,I.H., Spaink,H.P., Wijffelman,C.A. and
            Lugtenberg,B.J.
  TITLE     A Rhizobium leguminosarum biovar trifolii locus not localized on
            the sym plasmid hinders effective nodulation on plants of the pea
            cross-inoculation group
  JOURNAL   Mol. Plant Microbe Interact. 10 (7), 938-941 (1997)
  MEDLINE   97449849
   PUBMED   9304865
REFERENCE   2  (bases 1 to 33154)
  AUTHORS   Bladergroen,M.R., Badelt,K. and Spaink,H.P.
  TITLE     Infection-blocking genes of a symbiotic Rhizobium leguminosarum
            strain that are involved in temperature-dependent protein secretion
  JOURNAL   Mol. Plant Microbe Interact. 16 (1), 53-64 (2003)
  MEDLINE   22467694
   PUBMED   12580282
REFERENCE   3  (bases 1 to 33154)
  AUTHORS   Bladergroen,M.R., Badelt,K., Stronk,O.P., Lugtenberg,E.J.J. and
            Spaink,H.P.
  TITLE     An avirulence locus of a symbiotic Rhizobium leguminosarum strain
            is involved in temperature-dependent protein secretion
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 33154)
  AUTHORS   Bladergroen,M.R., Badelt,K., Stronk,O.P., Lugtenberg,E.J.J. and
            Spaink,H.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAR-2001) Institute of Molecular Plant Sciences,
            Leiden University, Wassenaarseweg 64, Leiden 2333 AL, The
            Netherlands
FEATURES             Location/Qualifiers
     source          1. .33154
                     /organism="Rhizobium leguminosarum bv. trifolii"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:386"
                     /note="biovar: trifolii"
     gene            744. .1820
                     /gene="impA"
     CDS             744. .1820
                     /gene="impA"
                     /function="involved in temperature-dependent protein
                     secretion"
                     /note="hypothetical"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="ImpA"
                     /protein_id="AAL17808.1"
                     /db_xref="GI:16326486"
                     /translation="MNVREIIDPLEQNHPCGENVRSNTAFREIYYRIKDARNAARTAE
                     RSIIPGETITIAPAWHDVSNLGLQLLSSKSKDIEVLAWLAEAQLRLRGFSGLHDVYVA
                     TVSLLDKHFDALHSIGDGDVEERFAPFAGLNGVGGEGTLIQAIRLTSLIPGGKFAQFS
                     LWDFQLSQRPNETERRKELQQAAAEAGVAQMSAYLDVLTGCIAAFDRMVEILDEHCGD
                     QAPPSSNTRNVLQEAASAIRMLAGIEAPTEAPLAAPQALASQGSEPLPASETDGSEAI
                     GLRPVSAEMIRSREEAFELLIAVARYFRRTEPHSPISMSIETLVRRGRMDFFELLAEL
                     LPEQQTRNAVLTAAGIQPVADKGG"
     gene            1857. .2402
                     /gene="impB"
     CDS             1857. .2402
                     /gene="impB"
                     /function="involved in temperature-dependent protein
                     secretion"
                     /note="hypothetical; similar to Edwardsiella ictaluri
                     hypothetical protein Eip20"
```

```
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="ImpB"
                     /protein_id="AAL17797.1"
                     /db_xref="GI:16326475"
                     /translation="MGRRRISMASVHEKLERVRKPRVHIKYEVETEGAMVVKELPFVV
                     GVLGDFSGNPTQPLKPFGERKFVQVDRDNFDDVMRRMTPGLTLSVENTLEKNGTELPV
                     SLKFESMADFEPGAVVDQVPGLKALLNARNQLRDLMSKADRSEELERLLEDILQNKAD
                     LSQLVAQLGGDKADEAKDALN"
     gene            2533. .4014
                     /gene="impC"
     CDS             2533. .4014
                     /gene="impC"
                     /function="involved in temperature-dependent protein
                     secretion"
                     /note="hypothetical; similar to Edwardsiella ictaluri
                     hypothetical protein Eip55"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="ImpC"
                     /protein_id="AAL17801.1"
                     /db_xref="GI:16326479"
                     /translation="MTAETQLQTKEDVGFAPDENLLSKVVAATRQTEPDRAQDLLRTL
                     TDQALKGTVTYDRNLTITLNNAIAEIDKAISGQLAAIMQSPEFTKLEGAWRGLNYLVK
                     NSETSANLKIRVLNASKRDVAKDLAKAVEFDQSRLFKSVYEDEFGTPGGEPMGALIGD
                     YEFDNSFDDVQLLQGVSMIAAAAFAPFISAASPRMFGFEDFRELAKPRDLEKIFETVE
                     YAKWRSLRDSDDSRFVTLAMPRVLARMPYGPKTSPIEEFNFDETGGSKTGELSHGSYC
                     WMNATYVMGTRLTEAFSKNGWCTAIRGAENGGKVEGLPMHIFSSDDGDLDLKCPTEVG
                     ITDRRDAELGKLGFLPLCHYKNTDYAVFFGAQTAHKPKLYDRPEATANAAVSARLPYM
                     MATSRFAHYLKVMGRDKIGSFMEAEDCEAWLNRWISNYVNANDDAGEESRAKYPLREA
                     KVTVQEIPGRPGAYNAVAWMRPWLQMEELTTSLRMVARIPSKN"
     gene            4081. .5469
                     /gene="impD"
     CDS             4081. .5469
                     /gene="impD"
                     /function="involved in temperature-dependent protein
                     secretion"
                     /note="hypothetical; similar to ImpC and to Edwardsiella
                     ictaluri hypothetical protein Eip55; possible duplication
                     of ImpC"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="ImpD"
                     /protein_id="AAL17802.1"
                     /db_xref="GI:16326480"
                     /translation="MGSSTEWRRRADQLITLIDEALNKQVNAILHHPDFQAMEARWRG
                     LAMLVREASRSGEVKVKLLGVTWSELTRTMERASDFDQSHLFELIYSREFGMPGGEPL
                     GLLVGDYYLSPADPAGGDPVSTLSQIGMVAAAAFCPFVAGAAPLAIGLEDFHELNRVN
                     DFSWLANETARIRWNGLRAREDSRFLGLVAPRILMRSPLTPHSRNRNDGFPFRESIAG
                     DGSSLLWGNAAFAFATVVIRNFIDSGWFADIRGVTQDAVDGGLLSTAELPPYDFGTES
                     NGLSAQAPVEIHLTSGQEQQLCELGIIPAATTYLSSSAIFNSNQSLHAPPHYSNEHAR
                     QNARLAAMLQYVLCASRFSHYLKVIMRDEIGQLADAISIERKLADWLTAYTLGNDDAD
                     IALRTRFPLRSAGIGVAEIPGKPGSFSCTVRLQPHFQLDDVATSFHLIAETATTGQIL
                     SRPVAIPERMSA"
     gene            5466. .6296
                     /gene="impE"
     CDS             5466. .6296
                     /gene="impE"
                     /function="involved in temperature-dependent protein
                     secretion"
                     /note="hypothetical"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="ImpE"
                     /protein_id="AAL17803.1"
                     /db_xref="GI:16326481"
                     /translation="MTLSASIAQSLSDNALDDAIEEVKAHLKVKPSDQEARHLYIDLL
                     VLAGDYQRADNQCSLAATLSPDATMGFALLRNELRAMAARDAWFTSGALPEFPQGPSE
                     LDKLAVRLGIAHRDGNADEARTTLDALEKLRGERPLIWNGKAVSDFRDLDDRTPHALE
                     VIMTGGGYLWIDFAKIAALSIEPIARPRDLAFRRAELSLIDGAAASVLLPAVYHGTGK
                     DETLRLGRETEWIEETTGITSGRGQRCYLAGDEMVSFHDTQSLEIVPAAAAGRQVAHG
                     "
     gene            6289. .6798
                     /gene="impF"
     CDS             6289. .6798
                     /gene="impF"
                     /function="involved in temperature-dependent protein
                     secretion"
                     /note="hypothetical; similar to Edwardsiella ictaluri
                     hypothetical protein Eip19"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="ImpF"
                     /protein_id="AAL17804.1"
                     /db_xref="GI:16326482"
                     /translation="MVDPLERYRLRDRVLASSILDRLIDEAPDRTVDPPMSFVDQVRD
                     VREAIRRDLEALLNTRRCPATPPAALSELRDALVSYGVDGVVSANLMTDQAKLKLAQA
                     IERRIALFETRLSDVRVTILKSRTMTERALRMRIQATFRLHEGMPPISFESTIDPSTQ
                     RFLVEAAHG"
     gene            6791. .8572
                     /gene="impG"
     CDS             6791. .8572
                     /gene="impG"
                     /function="involved in temperature-dependent protein
                     secretion"
                     /note="hypothetical"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="ImpG"
                     /protein_id="AAL17805.1"
                     /db_xref="GI:16326483"
                     /translation="MAEGFLHRYNEELLALRRRAARFADAFPKIAGRLRMTGDVADDP
                     HVERLIQSFAYSAARVRQKLDDEFPELTDGLLETLYPHYLAPLPSMSIVKFRPADTLA
                     SVQTVPRHTEVLAEPVRGEACRFRTTQDVEIAPIEITGANLSGQPINAPLSPYTGVAG
                     CLRLSIRSLDARKPLGGLGMSRLRLHIPSPWQQAVAIYELLANHTLGIALAKHADDRS
                     PVFLPAGNLKPVGFEPEEAMLPYPAPSFIGYRLLTEFFALPRKFLFLDIEGFDAWRGE
                     TLELFVYLGESDAKLERAVSVRDFALNATPVVNLFRQVCEPIAIDGTRTEYRLLPDAR
                     RQKTREIYSIDRVLLTGSRGQEEFCQPFFGRSQRAGTSATYWQAYRRFDEDDRTSDMD
                     IAFVNRQRRSSGPVDAVASVDTLCLNRDLPEQLPFGGGHPFLQLASGHEAIASVEALI
                     PPTPSVRMNDQEGRNWRLMSHLILNHLSLFDNDGAALKDILSLYAFRDSPETKQFVDA
                     LVRIKAGTSTARLGSGGTVPGTEIALEFDPAAIDRPSAFLFGSVLDRFFSLYASVNSF
                     TRLTIRMKGQSRPIARWPARAAERPLL"
     gene            8593. .9594
                     /gene="impH"
     CDS             8593. .9594
                     /gene="impH"
                     /function="involved in temperature-dependent protein
                     secretion"
                     /note="hypothetical"
                     /codon_start=1
                     /evidence=not_experimental

  Query Match             2.2%;  Score 80;  DB 1;  Length 33154;
  Best Local Similarity   54.3%;  Pred. No. 2.5e-06;
  Matches  183;  Conservative    0;  Mismatches  151;  Indels    3;  Gaps    1;

Qy       3252 AACATAAGCGCGTTAGTAAGAACCGTGTCAGCATCACCTATGACGTTGAAACGAATGGCG 3311
              |||    | || ||| | ||     || |||   |||||   || || || |||||   | ||||
Db       1897 AACTGGAACGGGTTCGCAAACCGCGAGTCCACATCAAATACGAAGTGGAAACCGAAGGCG 1956

Qy       3312 CCGTAAAGACGAAAGAGCTGCCGTTTGTTGTTGGCGTCATTGGCGACTTTTCAGGACACA 3371
              |  |    |  ||| ||  |||||  ||  ||  || ||  |||||   ||||||||  || ||| |
Db       1957 CGATGGTGGTGAAGGAACTGCCCTTCGTCGTCGGCGTGCTCGGCGACTTCTCCGGAAATC 2016

Qy       3372 AACCAGAATCAGAAAAAGTTGATTTAGAAGAGCGAGAGTTCACGGGTATCGATAAAGACA 3431
                 |    |     |     ||        ||| |   |||||  |||||   |     ||||     ||||
Db       2017 CCACGCAGCCGCTCAA---GCCTTTCGGCGAGCGCAAGTTCGTGCAGGTCGACCGCGACA 2073

Qy       3432 ACTTCGATACAGTGATGGGGCAAATTCACCCGCGTCTTTCGTACAAGGTTGATAACAAGC 3491
              |||||||     || |||  |  |    ||      ||| |  ||    |             ||| || ||||    |
```

```
Db       2074 ACTTCGACGACGTCATGCGCCGCATGACGCCGGGCCTCACCCTGTCGGTCGAGAACACTC 2133

Qy       3492 TTGCTAATGATGATAGCCAGTTTGAAGTGAACTTGAGCCTCCGTTCGATGAAAGATTTCC 3551

Db       2134 TCGAGAAGAACGGAACAGAGCTTCCCGTTTCGCTGAAATTCGAAAGCATGGCCGATTTCG 2193

Qy       3552 ACCCAGAGAACTTAGTTGATNAAATTGAGCCGCTTAA 3588

Db       2194 AACCCGGCGCGGTCGTCGATCAGGTGCCGGGGCTGAA 2230
```

RESULT 8
AE001398
LOCUS AE001398 14867 bp DNA linear INV 04-OCT-2002
DEFINITION Plasmodium falciparum 3D7 chromosome 2 section 35 of 73 of the complete sequence.
ACCESSION AE001398 AE001362
VERSION AE001398.1 GI:3845197
KEYWORDS .
SOURCE Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 14867)
AUTHORS Gardner,M.J., Tettelin,H., Carucci,D.J., Cummings,L.M., Aravind,L., Koonin,E.V., Shallom,S., Mason,T., Yu,K., Fujii,C., Pederson,J., Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Pertea,M., Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O., Smith,H.O., Fraser,C.M., Adams,M.D., Venter,J.C. and Hoffman,S.L.
TITLE Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum
JOURNAL Science 282 (5391), 1126-1132 (1998)
MEDLINE 99021743
PUBMED 9804551
REFERENCE 2 (bases 1 to 14867)
AUTHORS Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W., Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T., James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A., Kyes,S., Chan,M.-S., Nene,V., Shallom,S.J., Suh,B., Peterson,J., Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D., Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H., Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I., Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C., Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and Barrell,B.
TITLE Genome sequence of the human malaria parasite Plasmodium falciparum
JOURNAL Nature 419, 498-511 (2002)
REFERENCE 3 (bases 1 to 14867)
AUTHORS Gardner,M.J.
TITLE Direct Submission
JOURNAL Submitted (02-OCT-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES Location/Qualifiers

```
     source          1. .14867
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="2"
     repeat_region   complement(<1. .20)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(65. .137)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(178. .211)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   307. .327
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   complement(328. .348)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(378. .423)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(479. .503)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(513. .539)
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   577. .633
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   complement(714. .747)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   841. .867
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   complement(888. .918)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(1006. .1049)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(1075. .1120)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(1140. .1283)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   1294. .1346
                     /rpt_type=tandem
     repeat_region   1298. .1346
                     /rpt_type=tandem
                     /rpt_unit="(TATAA)n"
     repeat_region   complement(1423. .1531)
                     /note="AT_rich"
                     /rpt_type=tandem
     gene            complement(<1570. .>2424)
                     /locus_tag="PFB0490c"
     mRNA            complement(<1570. .>2424)
                     /locus_tag="PFB0490c"
     CDS             complement(1570. .2424)
                     /locus_tag="PFB0490c"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAC71887.1"
                     /db_xref="GI:3845198"
                     /translation="MKEKNEKIMDYLSCPLDDVVDREKKSGKNSLLKSSSTKKSDYKK
                     SSIFSKKRDSHKKGSSFRGRRSGFINRKSGSFKKPYYNNRLINKNYNNYKGRNFHNGR
                     DNFKGRTGSFGSRVFDNRKGSFKKRFISNRNKSSVKSYRGNGSNKMGRKSFNKAPTSR
                     TVVTKRLNNYKTVSAPVKKFNNLNISLYRKNRTFALNTKRSKPVGTIKSSVPRKRIKK
                     GLKKGSLKSKTRKSTSGSKFKPLNKYFLSKIKIVTSLNKIPSPLKEQKNTEVNLPESL
                     NNATTKKN"
     repeat_region   complement(2150. .2188)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(2446. .2468)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   2504. .2656
                     /rpt_type=tandem
                     /rpt_unit="(TATAA)n"
     repeat_region   2684. .2739
                     /rpt_type=tandem
     repeat_region   2693. .2736
                     /rpt_type=tandem
     repeat_region   complement(2697. .2732)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(2757. .2845)
                     /note="AT_rich"
                     /rpt_type=tandem
```

```
repeat_region   complement(2856. .2880)
                /rpt_type=tandem
                /rpt_unit="(TA)n"
repeat_region   2884. .2953
                /rpt_type=tandem
                /rpt_unit="(TATAA)n"
repeat_region   2888. .2924
                /rpt_type=tandem
repeat_region   complement(2977. .3167)
                /rpt_type=tandem
                /rpt_unit="(TAAA)n"
repeat_region   3047. .3132
                /rpt_type=tandem
repeat_region   complement(3047. .3132)
                /rpt_type=tandem
                /rpt_unit="(TA)n"
repeat_region   complement(3258. .3420)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(3450. .3514)
                /rpt_type=tandem
                /rpt_unit="(TA)n"
repeat_region   complement(3532. .3559)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(3689. .3720)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(4024. .4114)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   4107. .4146
                /rpt_type=tandem
repeat_region   complement(4115. .4146)
                /rpt_type=tandem
                /rpt_unit="(TA)n"
repeat_region   4227. .4261
                /rpt_type=tandem
repeat_region   complement(4227. .4257)
                /rpt_type=tandem
                /rpt_unit="(TA)n"
repeat_region   complement(4348. .4386)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(4403. .4450)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(4545. .4785)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(4937. .5001)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(5102. .5191)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(5246. .5275)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(5437. .5484)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(5554. .5577)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(6302. .6350)
                /note="AT_rich"
```

Query Match 2.2%; Score 79.4; DB 3; Length 14867;
Best Local Similarity 47.1%; Pred. No. 3.4e-06;
Matches 317; Conservative 0; Mismatches 346; Indels 10; Gaps 2;

```
Qy      523 TTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGC 582
```

```
Db     8749 TTTACTTATTTTAATTTAATTATTTTATTTTATATTATTTATTTATTATTATTTTAATTAATT 8808

Qy      583 TGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTG 642

Db     8809 TATTTAATATATTAATTTATTTTTATTTTATTTATTTATTTTTATTTTATTTATTTATTTATT 8868

Qy      643 TATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCA 702

Db     8869 ATTTATTTTTATTTTATTTATTTAATTATTTTAATTAATTTATTTAATATATTAATTTAT 8928

Qy      703 AGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTAT 762

Db     8929 TTTTATTTTATTTATTTATTTAATATATTATTTATTTTTATTTTATTTATTTATTATTTA 8988

Qy      763 TTTTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGA 822

Db     8989 AATTTATTTATTTAATATATTAATTTATTTTATTTTATTTATTTAATTATTTTAATTAAT 9048

Qy      823 CAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTT 882

Db     9049 TTATTTAATATATTAATTTATTTTTATATTATTTATTTATTATTTAAAT-------TTAT 9101

Qy      883 TTATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTC 942

Db     9102 TTATTTTATATTTTAATTAAATTATTTATTTTATATTATTTAATTATTATTTTAATTTAT 9161

Qy      943 TTTTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTAT 1002

Db     9162 TTATTTTATATTATTTGTTTATTATTTTAGTTTATTTATTATTTTAATTAAATTATTTAT 9221

Qy     1003 AATTCACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAA 1062

Db     9222 TTTATATTATTTATATATTATTTTTTAATTTAATTTATTATATTAAATTTATTTAATATT 9281

Qy     1063 TATTAGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTTGAAATAACTCAAG 1122

Db     9282 TAATATTTTAATTTTATTTATTTAATTAATTAATTATTTATTTATTTTAATATAATTATT 9341

Qy     1123 GCCTCTAGCGTTTTCCTTTATCTTAAAATACAGGAAATAGCGATTGAAGTTAATTGACAC 1182

Db     9342 TTTAATAAATATTTATTATATATTAATTTA---AAAATATATTTTTAAATTTATTTTTAT 9398

Qy     1183 TTAAGCAAATAGT 1195

Db     9399 TTAATCTTATATT 9411
```

```
RESULT 9
AE014833
LOCUS       AE014833               252394 bp    DNA     linear   INV 07-OCT-2002
DEFINITION  Plasmodium falciparum 3D7 chromosome 10 section 5 of 7 of the
            complete sequence.
ACCESSION   AE014833 AE014185
VERSION     AE014833.1  GI:23495098
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 252394)
  AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
            Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
            James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
            Kyes,S., Chan,M.-S., Nene,V., Shallom,S.J., Suh,B., Peterson,J.,
            Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
            Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
            Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
            Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
            Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
            Barrell,B.
  TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
  JOURNAL   Nature 419 (6906), 498-511 (2002)
   PUBMED   12368864
```

```
REFERENCE   2  (bases 1 to 252394)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .252394
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="10"
     repeat_region   complement(79. .213)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   253. .301
                     /rpt_type=tandem
     repeat_region   complement(253. .301)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   318. .374
                     /rpt_type=tandem
     repeat_region   324. .374
                     /rpt_type=tandem
                     /rpt_unit="(TAAA)n"
     repeat_region   complement(433. .482)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   577. .605
                     /rpt_type=tandem
     repeat_region   complement(577. .605)
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   complement(661. .682)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(792. .942)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   897. .1030
                     /rpt_type=tandem
                     /rpt_unit="(TATAA)n"
     repeat_region   complement(1038. .1173)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   1100. .1130
                     /rpt_type=tandem
     repeat_region   complement(1100. .1126)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   1133. .1172
                     /rpt_type=tandem
     repeat_region   1175. .1207
                     /rpt_type=tandem
     repeat_region   complement(1175. .1207)
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   complement(1218. .1281)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   1365. .1391
                     /rpt_type=tandem
     repeat_region   complement(1365. .1391)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(1401. .1493)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   1505. .1526
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   complement(1602. .1685)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(1726. .1825)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(1963. .2007)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(2049. .2157)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   2064. .2131
                     /rpt_type=tandem
     repeat_region   complement(2194. .2390)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   2257. .2332
                     /rpt_type=tandem
     repeat_region   complement(2405. .2457)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(2481. .2516)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   2532. .2615
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   complement(2623. .2657)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   2666. .2699
                     /rpt_type=tandem
                     /rpt_unit="(TAAA)n"
     repeat_region   complement(2740. .2807)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   2855. .3047
                     /rpt_type=tandem
     repeat_region   complement(2856. .3047)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   2893. .3047
                     /rpt_type=tandem
     repeat_region   3100. .3137
                     /rpt_type=tandem
     repeat_region   complement(3100. .3137)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(3145. .3170)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(3196. .3240)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   3283. .3319
                     /rpt_type=tandem
     repeat_region   complement(3287. .3319)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(3367. .3434)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(3513. .3656)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(3659. .3794)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(3809. .3838)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   3840. .3861
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
```

```
     repeat_region   complement(3866. .4026)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(4089. .4110)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   4111. .4295
                     /rpt_type=tandem
     repeat_region   complement(4111. .4235)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(4251. .4414)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(4262. .4295)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(4361. .4384)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(4424. .4483)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(4511. .4553)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(4602. .4628)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(4663. .4725)
                     /note="AT_rich"

  Query Match             2.1%;  Score 76.8;  DB 3;  Length 252394;
  Best Local Similarity   48.4%;  Pred. No. 1.1e-05;
  Matches  271;  Conservative    0;  Mismatches  287;  Indels    2;  Gaps    2;

Qy        541 ATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGCTGTTTATTGGATGAATAT 600
Db     147002 AAAATAAAAAAAAAAAAAAAAAAATTAAATATTACAAAATATATATTTTATATATTAATAT 147061

Qy        601 TCGAAATTATCACATAATAATTGATGCTATTATT-ACTTGCTGTATTGGTATCAACTTTC 659
Db     147062 ATAAATATATTAAATAAGATTTAGTTTTAATATTAATTTATTCTTATTTTATTATATATT 147121

Qy        660 ATGCTCTATACATGTA-ATATATTTCGAGTTAGACCTTAATTCAAGGTAATTTGTCTATT 718
Db     147122 TATATATATATATATATATATATATATTATATATATATTTTTTTTTTTTTTTCCTCTT 147181

Qy        719 TAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTTTATGTTTGTTTCA 778
Db     147182 TTATTTTTTATTATATTTTCAATTTTTTTTTTTTTTTTGTTTTGTTTTTCGTTTCTTAAT 147241

Qy        779 TTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAACATCTTTGATGAA 838
Db     147242 ATTATATTATAAGAAAAAAAATAAAAAAAATTATATCTTTATAAATATATATATAATAAT 147301

Qy        839 GTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTTATATTAAATCATAA 898
Db     147302 ATTTTTAATGTACTGTAAAAAATACATTAATTTCATAAAAAATGTATAATTATATTATAT 147361

Qy        899 TAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTTAAAGAGGTAGGA 958
Db     147362 ATAAATGAAAGTTTTTCTTTTTTTTTTTTTTTTTTTTGGGTATATATATATATATATAT 147421

Qy        959 TGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAATTCACCTCAAAGAT 1018
Db     147422 ATATATATATATATATATGTAAATTTATAATATTATTGTAATTTTTTTTCTTTAATCAT 147481

Qy       1019 ACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATATTAGTAGATGTTGT 1078
Db     147482 ATATAAATATGTAATATACATTATAAAATATATATATAATAATATATAAATATATAAATA 147541

Qy       1079 TTTTGCTGCAAAAATAAAAA 1098
```

```
Db     147542 TATATCTATAATAATAATAA 147561


RESULT 10
AP003630/c
LOCUS       AP003630              179902 bp    DNA     linear   HTG 21-MAR-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 6 clone P0566A10,
            *** SEQUENCING IN PROGRESS ***.
ACCESSION   AP003630
VERSION     AP003630.1  GI:14141691
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
            clone:P0566A10
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 179902)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .179902
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="6"
                     /clone="P0566A10"
ORIGIN

  Query Match             2.1%;  Score 76.2;  DB 2;  Length 179902;
  Best Local Similarity   48.5%;  Pred. No. 1.5e-05;
  Matches  210;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy        514 GTTTGGTGCTTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTT 573
Db      63539 GTTTTTTTTTTTTTTTTTTTATTTCTTTTTTTTTTTTTATTTTCTTTGTTATTTATTGTGT 63480

Qy        574 CTAAAAGGCTGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTAT 633
Db      63479 TTTTTTTTTTTTTTTTTGTGTATTTTTTTTTTTTTTTTAAATTCTTTTTTTTTTTTTT 63420

Qy        634 TACTTGCTGTATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGAC 693
Db      63419 TGTTTTTATTTTTTTATTTTTTTTTTTTTTTATTTGTTTTTTATATATTTGGTTTTTTTT 63360

Qy        694 CTTAATTCAAGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTT 753
Db      63359 TTTTTTTTTATTTTTTGTTTTTTATTTTTTTTTGTTTTATTTATTATTAATTTTTTATGT 63300

Qy        754 TGTGGTTATTTTTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTT 813
Db      63299 TTTTGTTTTCTATGTTGTTTTTTTATTTTTTCTTTTTTATTTTTTTTTTTATTCTTTTTTT 63240
```

```
Qy        814 TTATGATGACAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGC 873
              | |  |   |   | |||  |    | |||    | |  ||  |   |       |||
Db      63239 ATTTTTTCTTATATTTTTTTTTTTTTTTTTTTTATTTTCCTTTTTTTTTTTTAATTTTTA 63180

Qy        874 GTGTATTTTTTATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTT 933
              | | |||||| | |||| |  | |  | ||  | | | | | || || |   | |||||
Db      63179 TTATTTTTTTTTTTTTAATTTTTTAATAGATTGATATTTTTTTTTGTTATTCTTATTTTT 63120

Qy        934 ATAGTGTTCTTTT 946
              |  |||| ||||
Db      63119 TTTTTGTTTTTTT 63107
```

RESULT 11
AL162497/c

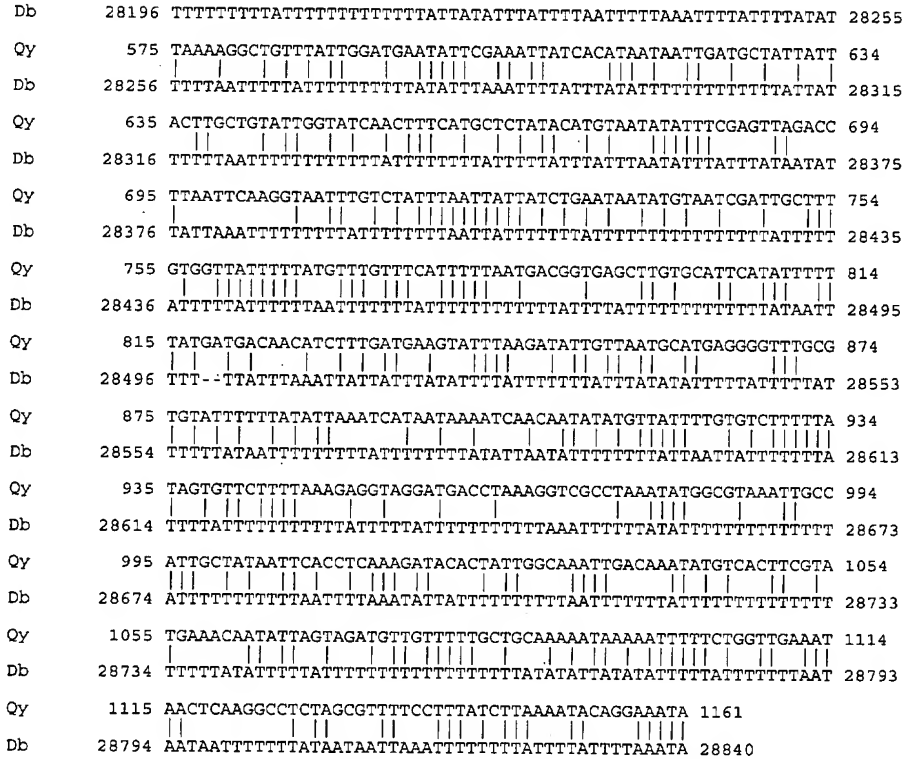
```
LOCUS       AL162497              143409 bp    DNA     linear   PRI 06-JUN-2001
DEFINITION  Human DNA sequence from clone RP11-313L9 on chromosome 13, complete
            sequence.
ACCESSION   AL162497
VERSION     AL162497.20  GI:14329908
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143409)
  AUTHORS   Mashreghi-Mohammadi,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 8, 2001 this sequence version replaced gi:14280409.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-313L9 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-313L9 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true right end of clone RP11-313L9 is at 143409 in this
            sequence. The true left end of clone RP11-40E6 is at 100074 in this
            sequence. The true right end of clone RP11-358F13 is at 100 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .143409
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-313L9"
                     /clone_lib="RPCI-11.2"
     repeat_region   1. .350
                     /note="L1MA9 repeat: matches 5421. .5789 of consensus"
     repeat_region   351. .654
                     /note="L1PB1 repeat: matches 5846. .6155 of consensus"
     repeat_region   655. .957
                     /note="L1MA9 repeat: matches 5789. .6085 of consensus"
     repeat_region   2061. .2106
                     /note="23 copies 2 mer tg 91% conserved"
     repeat_region   2063. .2106
                     /note="11 copies 4 mer tgtg 93% conserved"
     repeat_region   2138. .2422
                     /note="L1MD3 repeat: matches 6691. .7023 of consensus"
     repeat_region   2422. .3051
                     /note="L1MC4 repeat: matches 7168. .7837 of consensus"
     repeat_region   3674. .4545
                     /note="L1PA4 repeat: matches 5272. .6144 of consensus"
     repeat_region   5701. .5910
                     /note="7 copies 30 mer 89% conserved"
     repeat_region   6490. .6613
                     /note="MIR repeat: matches 92. .225 of consensus"
     repeat_region   9599. .9648
                     /note="25 copies 2 mer aa 94% conserved"
     repeat_region   10251. .10559
                     /note="AluSx repeat: matches 2. .312 of consensus"
     repeat_region   11134. .11521
                     /note="L1MB7 repeat: matches 5762. .6161 of consensus"
     repeat_region   12667. .12710
                     /note="11 copies 4 mer tggt 81% conserved"
     repeat_region   13210. .13309
                     /note="L2 repeat: matches 2559. .2674 of consensus"
     repeat_region   13794. .14148
                     /note="THE1C repeat: matches 1. .371 of consensus"
     repeat_region   14816. .15030
                     /note="MIR repeat: matches 49. .256 of consensus"
     repeat_region   14982. .15037
                     /note="L2 repeat: matches 2651. .2705 of consensus"
     repeat_region   15476. .15940
                     /note="L2 repeat: matches 1314. .1835 of consensus"
     repeat_region   16305. .16609
                     /note="AluSx repeat: matches 1. .300 of consensus"
     repeat_region   16683. .16722
                     /note="20 copies 2 mer tt 77% conserved"
     repeat_region   16948. .17073
                     /note="MLT1C repeat: matches 348. .464 of consensus"
     repeat_region   17074. .17360
                     /note="AluJo repeat: matches 13. .299 of consensus"
     repeat_region   17361. .17661
                     /note="MLT1C repeat: matches 7. .348 of consensus"
     repeat_region   17974. .18347
                     /note="L1MD1 repeat: matches 5659. .6029 of consensus"
     repeat_region   18399. .18554
                     /note="3 copies 52 mer 75% conserved"
     repeat_region   18400. .18549
                     /note="5 copies 30 mer 73% conserved"
     repeat_region   18413. .18552
                     /note="5 copies 28 mer 74% conserved"
     repeat_region   18565. .18738
                     /note="L1MD2 repeat: matches 6150. .6335 of consensus"
     repeat_region   19070. .19105
                     /note="9 copies 4 mer gata 88% conserved"
     repeat_region   19363. .19664
                     /note="L2 repeat: matches 2436. .2750 of consensus"
     repeat_region   20032. .20053
                     /note="11 copies 2 mer aa 100% conserved"
     repeat_region   21272. .21554
                     /note="AluSq repeat: matches 1. .283 of consensus"
     repeat_region   22754. .22918
                     /note="3 copies 55 mer 77% conserved"
     repeat_region   26601. .26684
                     /note="L2 repeat: matches 2668. .2748 of consensus"
     repeat_region   26685. .26814
                     /note="AluJb repeat: matches 2. .131 of consensus"
```

```
repeat_region   26815. .27110
                /note="AluSq repeat: matches 1. .303 of consensus"
repeat_region   27111. .27285
                /note="AluJb repeat: matches 130. .307 of consensus"
repeat_region   27286. .27578
                /note="L2 repeat: matches 2359. .2668 of consensus"
repeat_region   29945. .30257
                /note="AluY repeat: matches 1. .306 of consensus"
repeat_region   30812. .31078
                /note="AluSx repeat: matches 1. .267 of consensus"
repeat_region   31887. .32091
                /note="MER58A repeat: matches 1. .222 of consensus"
repeat_region   32818. .32853
                /note="9 copies 4 mer acac 100% conserved"
repeat_region   32856. .32957
                /note="MER97c repeat: matches 185. .289 of consensus"
repeat_region   32958. .33406
                /note="MLT1C repeat: matches 1. .466 of consensus"
repeat_region   33407. .33867
                /note="MER97c repeat: matches 289. .726 of consensus"
repeat_region   33895. .34718
                /note="8 copies 103 mer 61% conserved"
repeat_region   33915. .34519
                /note="121 copies 5 mer atata 56% conserved"
repeat_region   33920. .34103
                /note="92 copies 2 mer at 57% conserved"
repeat_region   33924. .34553
                /note="21 copies 30 mer 61% conserved"
repeat_region   33925. .34730
                /note="13 copies 62 mer 61% conserved"
repeat_region   33940. .34709
                /note="14 copies 55 mer 61% conserved"
repeat_region   33954. .34213
                /note="13 copies 20 mer 70% conserved"
repeat_region   33999. .34698
                /note="25 copies 28 mer 58% conserved"
repeat_region   34014. .34650
                /note="13 copies 49 mer 62% conserved"
repeat_region   34071. .34642
                /note="11 copies 52 mer 62% conserved"
repeat_region   34300. .34709
                /note="205 copies 2 mer aa 56% conserved"
repeat_region   34390. .34689
                /note="15 copies 20 mer 60% conserved"
repeat_region   34488. .34699
                /note="53 copies 4 mer taaa 59% conserved"
repeat_region   34512. .34703
                /note="64 copies 3 mer ata 60% conserved"
repeat_region   34629. .34718
                /note="3 copies 30 mer 78% conserved"
repeat_region   35062. .35647
                /note="L1ME1 repeat: matches 5550. .6144 of consensus"
repeat_region   35694. .35885
                /note="L2 repeat: matches 2054. .2250 of consensus"
repeat_region   35882. .36229
                /note="L2 repeat: matches 1446. .1832 of consensus"
repeat_region   36377. .36484
                /note="MIR repeat: matches 48. .147 of consensus"
repeat_region   39838. .39877
                /note="20 copies 2 mer tt 82% conserved"
repeat_region   39841. .39875
                /note="7 copies 5 mer tgttt 85% conserved"
repeat_region   41466. .41514
                /note="L2 repeat: matches 2699. .2747 of consensus"
repeat_region   41792. .42143
                /note="THE1B repeat: matches 1. .364 of consensus"
repeat_region   43471. .43590
                /note="L2 repeat: matches 2629. .2750 of consensus"
repeat_region   44937. .45157
                /note="L2 repeat: matches 2259. .2489 of consensus"

Query Match             2.1%;  Score 75.4;  DB 9;  Length 143409;
Best Local Similarity   47.6%;  Pred. No. 2.3e-05;
Matches  313;  Conservative    0;  Mismatches  341;  Indels    3;  Gaps    3;

Qy        461 ATTGCGTGACAATATAACTTTTGCTATCTGAAAGCTTAATGGTGAGGGAGCGGGTTTGGT 520
Db      34685 ATTTTATTATAATATTTATATTTATATAATATATATTTATATTTATATAATATATATTTA 34626

Qy        521 GCTTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAG 580
Db      34625 TATTTATATAATATATATTTATATAATATATATTTATATTTATATAATATATATTTATAT 34566

Qy        581 GCTGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGC 640
Db      34565 AATATATATTTATATTTATATAATATATATTATATAATATATAATATATATTATATAATA 34506

Qy        641 TGTATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATT 700
Db      34505 TATATTTATATTTA-TATAATATATATTATATAATATATATTTATATTATATAATATATA 34447

Qy        701 CAAGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTT 760
Db      34446 TTATATAATATATATTTATATTTATATAATATATATTATATAATATATATTTATATTTAT 34387

Qy        761 ATTTT-TATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGA 819
Db      34386 ATTATATATATTTATATAATATATATTTATATTTATATAATATATATTTATATTTATATA 34327

Qy        820 TGACAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTAT 879
Db      34326 TTTATATTTATTTATATTTATATATATTATATATTAAATATATGTTATATATAACTATAT 34267

Qy        880 TTTTTATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTG 939
Db      34266 ATTATATATTAAATATATGTTATATATAACATATATATTAAATATATATTATATAACTAT 34207

Qy        940 TTCTTTTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCAT-TG 998
Db      34206 ATATTATGTATTAAATATATATATTATATAACTATATATTATGTATTAAATATATATATT 34147

Qy        999 CTATAATTCACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAA 1058
Db      34146 ATATATTGTATATTAAATATATATTATATACATAACTTATATATGATATATATATATTATAT 34087

Qy       1059 ACAATATTAGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTTGAAATA 1115
Db      34086 ATTATATATGAAATATATATTATATATAACTATATATTATATATTATATATGAAATA 34030

RESULT 12
BX005205
LOCUS       BX005205               164188 bp    DNA     linear   HTG 30-OCT-2003
DEFINITION  Danio rerio clone DKEY-210P14, WORKING DRAFT SEQUENCE, 4 unordered
            pieces.
ACCESSION   BX005205
VERSION     BX005205.10  GI:36788966
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 164188)
  AUTHORS   Smith,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 27, 2003 this sequence version replaced gi:33468698.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
```

```
            Center project name: zK210P14
            ------------ Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 163185 bases at least Q40
            Consensus quality: 163490 bases at least Q30
            Consensus quality: 163732 bases at least Q20
            Insert size: 163888; sum-of-contigs
            Insert size: 157144; 9.9% error; agarose-fp
            Quality coverage: 6.58x in Q20 bases; sum-of-contigs Quality
            coverage: 6.92x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    82331: contig of 82331 bp in length
            *    82332    82431: gap of 100 bp
            *    82432    85590: contig of 3159 bp in length
            *    85591    85690: gap of 100 bp
            *    85691   107541: contig of 21851 bp in length
            *   107542   107641: gap of 100 bp
            *   107642   164188: contig of 56547 bp in length.
FEATURES             Location/Qualifiers
     source          1. .164188
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-210P14"
                     /clone_lib="DanioKey"
     misc_feature    1. .82331
                     /note="assembly_fragment:01879
                     fragment_chain:1"
     misc_feature    82432. .85590
                     /note="assembly_fragment:00792
                     fragment_chain:1"
     misc_feature    85691. .107541
                     /note="assembly_fragment:01113
                     fragment_chain:1"
     misc_feature    107642. .164188
                     /note="assembly_fragment:01413
                     fragment_chain:1"
ORIGIN

  Query Match             2.1%;  Score 74.8;  DB 2;  Length 164188;
  Best Local Similarity   51.3%;  Pred. No. 3e-05;
  Matches  199;  Conservative    0;  Mismatches  187;  Indels    2;  Gaps    1;

Qy        558 TATTAAATAAACAGTTCTAAAAGGCTGTTTATTGGATGAATATTCGAAATTATCACATAA 617
              |||||    |    | ||  |   | ||||||||||   |||| |||| | |    ||
Db     135066 TTTTAAAATTATTTATTTATTAATTTATTTATTGGATTTTTATTGGAAAATGTATTTTAT 135125

Qy        618 TAATTGATGCTATTATTACTTGCTGTATTGGTATCAACTTTCATGCTCTATACATGTAAT 677
              | ||| ||  |||| |   |   |  |  |   ||| ||   |||    | ||
Db     135126 TTATTTATTTTATTTATTTGTTTATTTATTTGACTTTATTTTATTAGTTATTTGATTTAT 135185

Qy        678 ATATTTCGAGTTAGACCTTAATTCAAGGTAATTTGTCTATTTAATTATTATCTGAATAAT 737
               ||||      |  |  || |||  || |  |||| | |||| |  |||| |||||| |
Db     135186 TTATTCATTTATTTATTTTCATTTTAGTTTGTTTGTTTATTTGACTATTTATTGAATATT 135245

Qy        738 ATGTAATCGATTGCTTTGTGGTTATTTTTATGTTTGTTTCATTTTTAATGACGGTGAGCT 797
              | |   | |  ||| |  || |||||||   || ||||||||   ||      |    |
Db     135246 TATTTAATAAAT--TTTTTATTTGTTTTTACTGTTATTTCATTTGCTATTTGATTATTAT 135303

Qy        798 TGTGCATTCATATTTTTTATGATGACAACATCTTTGATGAAGTATTTAAGATATTGTTAA 857
              |      ||  | ||| ||    |||  |  | |||  | || | ||||     |   ||
Db     135304 TTATTTTTGTTGTTTATTTATTTGATTATTTATTTATTTAATTGTTTATTTATTAAAAAA 135363

Qy        858 TGCATGAGGGGTTTGCGTGTATTTTTTATATTAAATCATAATAAAATCAACAATATATGT 917
```

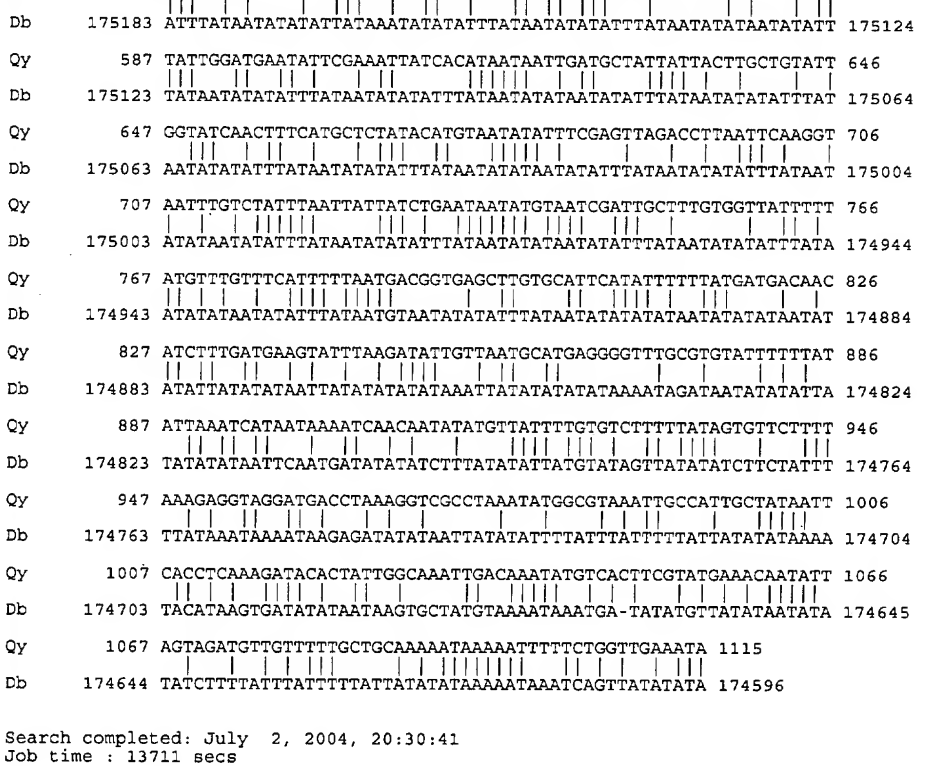
```
                  | || |  || |  ||| | |||| |||  |  |  | |  ||  |  || |
Db     135364 AATCATTTGAGTATTTGTTTTATTTATTTATTTAATATTTTTTTATTTGACTGTTAATTT 135423

Qy        918 TATTTTGTGTCTTTTTATAGTGTTCTTT 945
              |||||    | | |||||    || |||
Db     135424 TATTTGTGATTTATTTATTTATTTATTT 135451


RESULT 13
PFMAL7P1_07
WPCOMMENT
  Sequence split into 14 fragments  LOCUS PFMAL7P1 Accession AL844506
     Fragment Name      Begin     End
     PFMAL7P1_00            1    110000
     PFMAL7P1_01       100001    210000
     PFMAL7P1_02       200001    310000
     PFMAL7P1_03       300001    410000
     PFMAL7P1_04       400001    510000
     PFMAL7P1_05       500001    610000
     PFMAL7P1_06       600001    710000
     PFMAL7P1_07       700001    810000
     PFMAL7P1_08       800001    910000
     PFMAL7P1_09       900001   1010000
     PFMAL7P1_10      1000001   1110000
     PFMAL7P1_11      1100001   1210000
     PFMAL7P1_12      1200001   1310000
     PFMAL7P1_13      1300001   1351552
Continuation (8 of 14) of PFMAL7P1 from base 700001 (AL844506 Plasmodium falciparum 3D7

  Query Match             2.1%;  Score 74.6;  DB 2;  Length 110000;
  Best Local Similarity   48.2%;  Pred. No. 3.4e-05;
  Matches  279;  Conservative    0;  Mismatches  289;  Indels   11;  Gaps    2;

Qy        523 TTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGC 582
              | ||  |  || |  | || || ||||| || | || || ||| | |  | ||  |
Db      15055 TATATATATATATTTATTATTTTAAATTATTTAATTAATATATATATATTTATTATTT 15114

Qy        583 TGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTG 642
              |   |  |  |   |  | | |||  | || ||| |  ||  | | |||| |   ||
Db      15115 TAAATTATTTAATTAATATATATATTTAAATATTATATTAAATTATTTAATTAATATGTG 15174

Qy        643 TATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCA 702
              ||||    ||  ||| |   |  ||        |||| |  | |||    || | |
Db      15175 TATTTAATTCTTATTTTAAATTAATTA--------ATATATATATTTATTTATTCAATTG 15226

Qy        703 AGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTAT 762
              ||| || | || | ||| || || |  | | |  |  ||  |||     |||
Db      15227 TTTTAAATTATTTAATAAATAATAATATTGTTTTTTTTAATGAATTTAATTTTATTATAT 15286

Qy        763 TTTTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGA 822
              ||  | | |  |  | |  |  | | |  |  ||  | |||  | ||  |
Db      15287 ATTATTATATTATATTATATATTTTTTTTTAATAATTAATTTTATTTATTTATGTGTATA 15346

Qy        823 CAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTT 882
              | ||  ||    | ||  | | || | |  || |    |  |||||   |
Db      15347 TTTTTTTTTATATTCCATTTTACATTTTTTATAATTATATTATTTATAAATGTGTAAACT 15406

Qy        883 TTATATTAAATCATAATAAAATCAACAATATAT---GTTATTTTGTGTCTTTTTATAGTG 939
              |||||| |||   ||   | || || |||||||   ||  |  |  | |||||||| |
Db      15407 TTATATAAAAATTTATATATATTAATAATATATTTAGTAGTAATAATTATTTTTTATATTT 15466

Qy        940 TTCTTTTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGC 999
              || ||| |   |   | | ||||| ||      ||||||       ||  |  |  |
Db      15467 CATCTTAAAAAATTCATTTTTAACTAAATGTATTTAAAATATTTATATAAGAGATAAAGA 15526

Qy       1000 TATAATTCACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAA 1059
              ||| |||||  |  || |   |  | | |  ||  | || |  | |  || | |||
Db      15527 TATGATTCAAATGTAAAAATATATTATTGAACTTTATAAGATTTTATAATATGTTTTAAA 15586

Qy       1060 CAATATTAGTAGATGTTGTTTTTGCTGCAAAAATAAAAA 1098
              ||| | | ||| |   |  | || ||   | | ||| ||
Db      15587 CAAAAATTGTACAATATAATCTTCCTATTACAGTAATAA 15625
```

```
RESULT 14
AC125567
LOCUS       AC125567              170627 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-9H22, *** SEQUENCING IN PROGRESS ***,
            4 unordered pieces.
ACCESSION   AC125567
VERSION     AC125567.5  GI:24817949
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 170627)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwackelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170627)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 170627)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 9, 2002 this sequence version replaced gi:23096537.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDDV
                Center clone name: CH230-9H22
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 162619 bases at least Q40
                Consensus quality: 164392 bases at least Q30
                Consensus quality: 165269 bases at least Q20
                Estimated insert size: 169231; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   109158: contig of 109158 bp in length
            *  109159   109258: gap of unknown length
            *  109259   110280: contig of 1022 bp in length
            *  110281   110380: gap of unknown length
            *  110381   111540: contig of 1160 bp in length
            *  111541   111640: gap of unknown length
            *  111641   170627: contig of 58987 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170627
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-9H22"
     misc_feature    1. .1284
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    3636. .4924
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH341988"
ORIGIN

  Query Match             2.1%;  Score 74.6;  DB 2;  Length 170627;
  Best Local Similarity   45.7%;  Pred. No. 3.3e-05;
  Matches  296;  Conservative    0;  Mismatches  349;  Indels    2;  Gaps    1;

Qy        515 TTTGGTGCTTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTC 574
              ||| |  | |  ||  ||  |  | ||||| | || | |   ||    ||     |
```



RESULT 15
AC131649/c
LOCUS AC131649 182518 bp DNA linear PRI 26-NOV-2003
DEFINITION Homo sapiens chromosome 16 clone RP11-609N14, complete sequence.
ACCESSION AC131649 AC016139
VERSION AC131649.2 GI:38524629
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 182518)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 182518)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 182518)
AUTHORS Stanford Human Genome Center and Los Alamos National Laboratory.
CONSRTM DOE Joint Genome Institute
TITLE Direct Submission
JOURNAL Submitted (26-NOV-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Nov 26, 2003 this sequence version replaced gi:22475305.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos National Laboratory
www-shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
FEATURES Location/Qualifiers
source 1. .182518
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-609N14"
ORIGIN

Query Match 2.1%; Score 74.6; DB 9; Length 182518;
Best Local Similarity 46.5%; Pred. No. 3.3e-05;
Matches 274; Conservative 0; Mismatches 314; Indels 1; Gaps 1;



Search completed: July 2, 2004, 20:30:41
Job time : 13711 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 2, 2004, 14:05:42 ; Search time 1304 Seconds
(without alignments)
11689.060 Million cell updates/sec

Title: US-10-780-347-1
Perfect score: 3588
Sequence: 1 gtcgacttattgcattgatg.....gatnaaattgagccgcttaa 3588

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

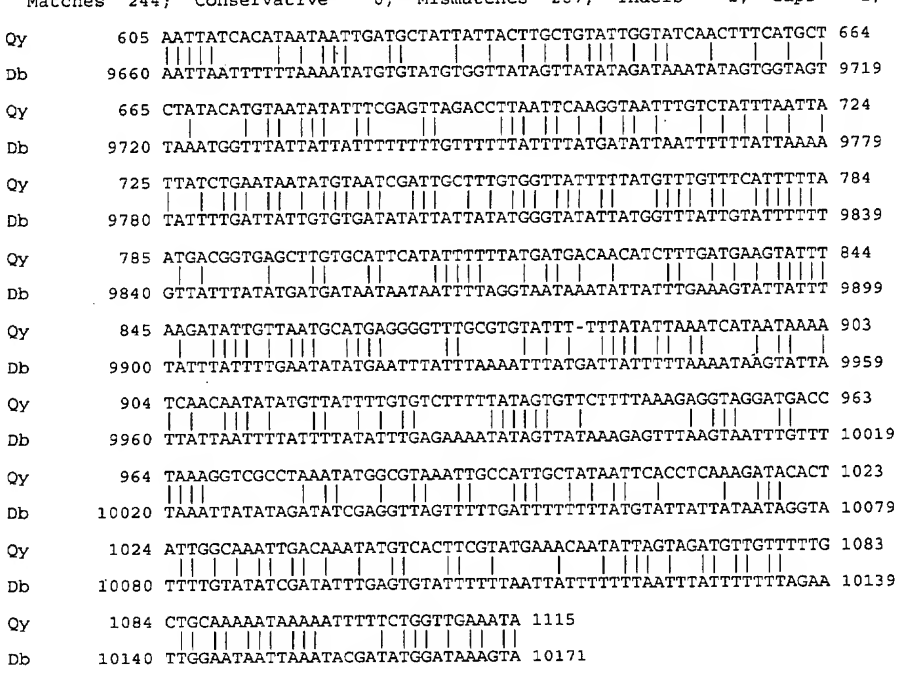
Result No.	Score	% Query Match	Length	DB	ID	Description
1	72.8	2.0	11422	6	ABK39937	Abk39937 Human che
2	72.8	2.0	11422	6	ABL32219	Abl32219 Human imm
3	72.4	2.0	8093	7	ABZ10178	Abz10178 Haematopo
4	72.4	2.0	8093	9	ADB54268	Adb54268 Pretreate
5	72.4	2.0	8093	9	ADE84190	Ade84190 Human lym
6	71.4	2.0	61020	4	AAS46788	Aas46788 Tumour su
7	69.8	1.9	8666	9	ADB54239	Adb54239 Pretreate
8	69.8	1.9	8666	9	ADE84177	Ade84177 Human lym
9	69.4	1.9	8093	4	AAS46436	Aas46436 Tumour su
10	69.4	1.9	8093	6	ABK33974	Abk33974 Human DNA
11	69.4	1.9	8093	6	ABL92237	Abl92237 Chemicall
12	69.4	1.9	8093	6	ABL49332	Abl49332 Human pol
13	69.4	1.9	8093	7	ABZ10032	Abz10032 Haematopo
14	69.4	1.9	8093	7	ADA20361	Ada20361 Prostate
15	69.4	1.9	8093	7	ADA84168	Ada84168 Human ren
16	69.4	1.9	8093	9	ADB54140	Adb54140 Pretreate
17	69.4	1.9	8093	9	ADE84114	Ade84114 Human lym
18	69	1.9	5504	6	ABL70350	Abl70350 Chemicall
19	69	1.9	5504	6	AAS61309	Aas61309 Human gen
20	68.2	1.9	83391	6	ABQ67094	Abq67094 Human ang
21	68	1.9	6306	4	AAS45516	Aas45516 Chemicall
22	68	1.9	6306	6	ABK28458	Abk28458 DNA trans
23	68	1.9	13131	6	ABL92249	Abl92249 Chemicall
c 24	66.8	1.9	8056	7	ABZ10246	Abz10246 Haematopo
25	66.6	1.9	8666	4	AAS46305	Aas46305 Tumour su
26	66.6	1.9	8666	6	ABL32396	Abl32396 Human imm
27	66.6	1.9	8666	6	ABK34008	Abk34008 Human DNA
28	66.6	1.9	8666	6	ABQ67177	Abq67177 Human ang
29	66.6	1.9	8666	9	ADB54111	Adb54111 Pretreate
30	66.6	1.9	8666	9	ADE84101	Ade84101 Human lym
31	66.6	1.9	9504	6	ABK28407	Abk28407 DNA trans
32	66.2	1.8	17848	4	AAS45323	Aas45323 Chemicall
33	66.2	1.8	17848	6	ABK39976	Abk39976 Human che
34	66.2	1.8	17848	6	ABK28164	Abk28164 DNA trans
35	66.2	1.8	18624	6	ABL33702	Abl33702 Human imm
36	66	1.8	516	7	ABX40620	Abx40620 Bovine ES
c 37	66	1.8	19124	2	AAT72882	Aat72882 Plasmodiu
c 38	66	1.8	19124	3	AAZ98287	Aaz98287 Plasmodiu
39	65.8	1.8	7849	6	ABL92279	Abl92279 Chemicall
40	65.8	1.8	7849	6	AAD22330	Aad22330 Chemicall
41	65.8	1.8	19459	6	ABK31212	Abk31212 Signal tr
42	65.8	1.8	19459	6	ABL70527	Abl70527 Chemicall
43	65.4	1.8	1179	6	ABQ49604	Abq49604 Oligonucl
c 44	65.4	1.8	1179	6	ABQ49605	Abq49605 Oligonucl
45	65.2	1.8	6040	6	ABL92300	Abl92300 Chemicall

ALIGNMENTS

RESULT 1
ABK39937
ID ABK39937 standard; DNA; 11422 BP.
XX
AC ABK39937;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human chemically pretreated gene sequence #9 strand 2.
XX
KW Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN WO200202806-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007470.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154757/20.
XX
PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT useful for detecting cytosine methylation state of genes associated with
PT pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS Claim 1; SEQ ID NO 18; 24pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes
CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996,
CC NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC their complementary sequences, or a sequence (S1) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite

CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 11422 BP; 3913 A; 59 C; 1856 G; 5594 T; 0 U; 0 Other;

Query Match 2.0%; Score 72.8; DB 6; Length 11422;
Best Local Similarity 47.7%; Pred. No. 1.2e-06;
Matches 244; Conservative 0; Mismatches 267; Indels 1; Gaps 1;



RESULT 2
ABL32219
ID ABL32219 standard; DNA; 11422 BP.
XX
AC ABL32219;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 192.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 192; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 11422 BP; 3913 A; 59 C; 1856 G; 5594 T; 0 U; 0 Other;

Query Match 2.0%; Score 72.8; DB 6; Length 11422;
Best Local Similarity 47.7%; Pred. No. 1.2e-06;
Matches 244; Conservative 0; Mismatches 267; Indels 1; Gaps 1;

```
Db      10080 TTTTGTATATCGATATTTGAGTGTATTTTTTAATTATTTTTTTAATTTATTTTTTTAGAA 10139

Qy       1084 CTGCAAAAATAAAAATTTTTCTGGTTGAAATA 1115
              || || ||| |||     | ||| | || ||
Db      10140 TTGGAATAATTAAATACGATATGGATAAAGTA 10171
```

RESULT 3
ABZ10178
ID ABZ10178 standard; DNA; 8093 BP.
XX
AC ABZ10178;
XX
DT 16-JAN-2003 (first entry)
XX
DE Haematopoietic cell proliferation disorder related DNA sequence #318.
XX
KW Human; haematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW cytosine methylation state; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277272-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-EP003401.
XX
PR 26-MAR-2001; 2001US-0278333P.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI Schwope I, Ziebarth H;
XX
DR WPI; 2003-018942/01.
XX
PT Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent that
PT distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS Claim 28; SEQ ID NO 318; 117pp; English.
XX
CC The present invention describes a method for detecting and
CC differentiating between haematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used: for
CC differentiating between healthy haematopoietic cells and proliferative
CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related DNA
CC sequences. The nucleotide sequences from the present invention can also
CC be used for detecting a predisposition to, differentiation between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC haematopoietic cell proliferative disorders. The present method enables a
CC highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients
XX
SQ Sequence 8093 BP; 2419 A; 0 C; 1689 G; 3985 T; 0 U; 0 Other;

Query Match 2.0%; Score 72.4; DB 7; Length 8093;
Best Local Similarity 47.6%; Pred. No. 1.4e-06;

Matches 283; Conservative 0; Mismatches 301; Indels 10; Gaps 2;

```
Qy       516 TTGGTGCTTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCT 575
             || |||      |||  |   |  |  ||   ||||   |    | ||||||    ||| |
Db      6274 TTTGTGTGGGTATTGTTGTTTTATTTTAAAAAATTTTGTTTAAAGTAAATATTTTGTTTT 6333

Qy       576 AAAAGGCTGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTA 635
             ||||  |  |  |     |        | ||  ||  |||  |    ||   |  |  |||   |
Db      6334 AAAATGATATAGTATTAGATTTTGTTAGATGTTAGAAATGGATTTATTTTAAAATTTGGA 6393

Qy       636 CTTGCTGTATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCT 695
             |||  |||||    |  |  |  |  |      ||||  |  |||   |   ||  |         |
Db      6394 ATTGTTGTATATATTTTATATGTAAGATAGTATATAAGTAGAAATATTTAAAAGTAGTTT 6453

Qy       696 TAATTCAAGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTG 755
             ||  ||   ||  |     |  ||  |||  ||||     |||||  |||   |    ||     |
Db      6454 TATTTATAGATTGTAGTAATTTTGTATTTTTATTAAGATAATTTGTTTTGTGTTAAAATA 6513

Qy       756 TGGTTATTTTTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTT 815
                 |  |||   ||  |||||||   |  |   ||  |    |          ||  |||          ||
Db      6514 GTAATTTTTAAATTTTTGTTTATTATGAAAAGGTAATTTTAAAGTTTATTATGTAAAATT 6573

Qy       816 ATGATGACAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGT 875
             |      |  |     |||  ||   |    ||   ||||  ||  ||||     |       |||
Db      6574 AATTATAAATAGGATTTAATTTATATTTATAGATTTTTTTAAGTATTATATAATTTAA-- 6631

Qy       876 GTATTTTTTATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTAT 935
             ||||||   |  |  ||  |   |   |||  |    | ||||  |||  ||||||   | ||  ||
Db      6632 AAATTTTTGTTTTAAAGTTTTGTTTTAATTATTGAAATATTTTAATTTGTGGTTATTAAT 6691

Qy       936 AGTGTTCTTTTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCA 995
                 ||   |||||  ||  ||||  |     | |||         |||  |  |     |||||
Db      6692 TTAGTAAATTTAAGGAATTAGGTTATGATTAAG--------AATTTAGGTGGAATTGATG 6743

Qy       996 TTGCTATAATTCACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTAT 1055
             ||    |||||  |  | ||  |  ||  |  |  |||  |  |  |  |  |||  |  |  |  | |
Db      6744 TTTGGGTAATTAAAATAAATGGTATAAGAGTTTAAAAATTAAAGTTGTGTTAGTGGTTTT 6803

Qy      1056 GAAACAATATTAGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTT 1109
             ||  |    |||||   |  |  ||||   |  |      | ||||  |  |  || |
Db      6804 TAATTAGAGGTAGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTGTAGGGGT 6857
```

RESULT 4
ADB54268
ID ADB54268 standard; DNA; 8093 BP.
XX
AC ADB54268;
XX
DT 04-DEC-2003 (first entry)
XX
DE Pretreated genomic DNA region 192.
XX
KW colon cell proliferative disorder; non methylated CpG dinucleotide;
KW cytostatic; cancer; adenoma; carcinoma; cytosine methylation state; ds.
XX
OS Unidentified.
XX
PN WO2003072821-A2.
XX
PD 04-SEP-2003.
XX
PF 27-FEB-2003; 2003WO-EP002035.
XX
PR 27-FEB-2002; 2002EP-00004551.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Adorjan P, Burger M, Maier S, Nimmrich I, Becker E, Lesche R;
PI Rujan T, Schmitt A;
XX
DR WPI; 2003-731620/69.
XX

```
PT   Detecting and differentiating between colon cell proliferative disorders
PT   associated with a gene or its regulatory regions comprises contacting a
PT   target nucleic acid in a biological sample obtained from the subject with
PT   a reagent.
XX
PS   Claim 32; SEQ ID NO 324; 74pp; English.
XX
CC   The invention relates to a novel method for detecting and differentiating
CC   between colon cell proliferative disorders associated with at least one
CC   gene or its regulatory regions. The method comprises contacting a target
CC   nucleic acid in a biological sample obtained from the subject with at
CC   least one reagent or a series of reagents, where the reagent or series of
CC   reagents, distinguishes between methylated and non methylated CpG
CC   dinucleotides within the target nucleic acid. The molecules of the
CC   invention demonstrate cytostatic activity whilst the method may useful
CC   for detecting and differentiating between colon cell proliferative
CC   disorders, including cancers such as colon adenoma and colon carcinoma.
CC   The PNA (peptide nucleic acid)-oligomers are useful as probes for
CC   determining cytosine methylation state or single nucleotide
CC   polymorphisms. The current sequence is that of the pretreated genomic DNA
CC   region of the invention. This sequence is not shown within the
CC   specification but is taken from Wipoweb.
XX
SQ   Sequence 8093 BP; 2419 A; 0 C; 1689 G; 3985 T; 0 U; 0 Other;

  Query Match             2.0%;  Score 72.4;  DB 9;  Length 8093;
  Best Local Similarity   47.6%;  Pred. No. 1.4e-06;
  Matches  283;  Conservative    0;  Mismatches  301;  Indels   10;  Gaps    2;

Qy        516 TTGGTGCTTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCT 575
              || |||     |||  |   |  |  ||   ||||    |     | ||||||    ||| |
Db       6274 TTTGTGTGGGTATTGTTGTTTTATTTTAAAAAATTTTGTTTAAAGTAAATATTTTGTTTT 6333

Qy        576 AAAAGGCTGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTA 635
              ||||  | | |    |       | ||  ||  |||  |    ||   |  |   |||   |
Db       6334 AAAATGATATAGTATTAGATTTTGTTAGATGTTAGAAATGGATTTATTTTAAAATTTGGA 6393

Qy        636 CTTGCTGTATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCT 695
               |||  |||||   |  |  |  | |     ||||  | |||  |   ||  |       |
Db       6394 ATTGTTGTATATATTTTATATGTAAGATAGTATATAAGTAGAAATATTTAAAAGTAGTTT 6453

Qy        696 TAATTCAAGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTG 755
              || ||  || |  |      | ||  ||| ||||    |||||  |||  |   ||    |
Db       6454 TATTTATAGATTGTAGTAATTTTGTATTTTTATTAAGATAATTTGTTTTGTGTTAAAATA 6513

Qy        756 TGGTTATTTTTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTT 815
                | |||  ||  |||||||  | |  || |    |        ||  |||         ||
Db       6514 GTAATTTTTAAATTTTTGTTTATTATTATGAAAAGGTAATTTTAAAGTTTATTATGTAAAATT 6573

Qy        816 ATGATGACAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGT 875
              |     | |    |||  ||  |   ||  ||||  ||  ||||     |      |||
Db       6574 AATTATAAATAGGATTTAATTTATATTTATAGATTTTTTTAAGTATTATATAATTTAA-- 6631

Qy        876 GTATTTTTTATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTAT 935
              ||||||  | | ||  |   |  |  |||  |    | ||||  |||  ||||||   | ||  ||
Db       6632 AAATTTTTGTTTTAAAGTTTTGTTTTAATTATTGAAATATTTTAATTTGTGGTTATTAAT 6691

Qy        936 AGTGTTCTTTTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCA 995
                ||   |||||  ||  ||||  |     | |||         |||  |  |    |||||
Db       6692 TTAGTAAATTTAAGGAATTAGGTTATGATTAAG--------AATTTAGGTGGAATTGATG 6743

Qy        996 TTGCTATAATTCACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTAT 1055
              ||    |||||  |  | ||  | ||  |  |  ||| | |  |  | ||| |  |   | |
Db       6744 TTTGGGTAATTAAAATAAATGGTATAAGAGTTTAAAAATTAAAGTTGTGTTAGTGGTTTT 6803

Qy       1056 GAAACAATATTAGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTT 1109
               ||  |    |||||  |  |  ||||   | |      | ||||  |  |  || |
Db       6804 TAATTAGAGGTAGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTGTAGGGGT 6857


RESULT 5
ADE84190
ID   ADE84190 standard; DNA; 8093 BP.
XX
AC   ADE84190;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human lymphoid cell proliferative disorder gene derived DNA #126.
XX
KW   ds; lymphoid cell proliferative disorder; methylation;
KW   methylated CpG dinucleotide; single nucleotide polymorphism; SNP;
KW   diffuse large B-cell lymphoma; mantle cell lymphoma;
KW   chronic lymphocytic leukemia; small lymphocytic lymphoma;
KW   follicular lymphoma; diagnosis; prognosis.
XX
OS   Homo sapiens.
XX
PN   WO2003044226-A2.
XX
PD   30-MAY-2003.
XX
PF   25-NOV-2002; 2002WO-EP013265.
XX
PR   23-NOV-2001; 2001DE-01057491.
PR   28-DEC-2001; 2001DE-01064501.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Burger M,  Caldwell C,  Genc B,  Becker E,  Maier S,  Nimmrich I;
XX
DR   WPI; 2003-457621/43.
XX
PT   Detecting and differentiating between lymphoid cell proliferative
PT   disorders comprises contacting a target nucleic acid with at least one
PT   reagent that distinguishes between methylated and non-methylated CpG
PT   dinucleotides.
XX
PS   Claim 26; SEQ ID NO 186; 448pp; English.
XX
CC   The invention relates to a method of detecting and differentiating
CC   between lymphoid cell proliferative disorders associated with at least
CC   one gene and/or their regulatory regions in a subject by contacting a
CC   target nucleic acid in a biological sample obtained from the subject with
CC   at least one reagent or series of reagents that distinguish between
CC   methylated and non-methylated CpG dinucleotides within the target nucleic
CC   acid. The genes and/or their regulatory regions are preferably selected
CC   from MDR1, CSNK2B, EGR4, AR, CDK4, RB2, CDC25A, GPIb beta, MYOD1, CDH3,
CC   MYCL1, ELK1, ABL1, APC, BCL2, CDH1, CDKN1A, CDKN1B, CDKN2a, CDKN2B, FOS,
CC   GSTP1, HIC-1, MGMT, MLH1, MOS, MYC, PTEN, RBL2, TGFBR2, TP73, CDKNIC,
CC   GSK3beta, ESR1, APAF1, BAK1, BAX or HOXA5. Oligomers, peptide nucleic
CC   acid (PNA)-oligomers and/or isolated nucleic acids based on the sequences
CC   of the genes are useful for detecting the methylation state of all the
CC   CpG dinucleotides within one or more the sequences, or their complements,
CC   for determining the cytosine methylation state and or single nucleotide
CC   polymorphisms (SNPs), and for differentiating at least two of the medical
CC   conditions such as diffuse large B-cell lymphoma, mantle cell lymphoma,
CC   chronic lymphocytic leukemia, small lymphocytic lymphoma and follicular
CC   lymphoma. They are also useful for detecting of a predisposition to,
CC   differentiation between subclasses, diagnosis, prognosis, treating and/or
CC   monitoring of lymphoid cell proliferative disorder. This sequence
CC   represents a nucleic acid of a pretreated genomic DNA derived from the
CC   above mentioned genes.
XX
SQ   Sequence 8093 BP; 2419 A; 0 C; 1689 G; 3985 T; 0 U; 0 Other;

  Query Match             2.0%;  Score 72.4;  DB 9;  Length 8093;
  Best Local Similarity   47.6%;  Pred. No. 1.4e-06;
  Matches  283;  Conservative    0;  Mismatches  301;  Indels   10;  Gaps    2;

Qy        516 TTGGTGCTTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCT 575
              || |||     |||  |   |  |  ||   ||||    |     | ||||||    ||| |
Db       6274 TTTGTGTGGGTATTGTTGTTTTATTTTAAAAAATTTTGTTTAAAGTAAATATTTTGTTTT 6333

Qy        576 AAAAGGCTGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTA 635
              ||||  | | |    |       | ||  ||  |||  |    ||   |  |   |||   |
```

```
Db       6334 AAAATGATATAGTATTAGATTTTGTTAGATGTTAGAAATGGATTTATTTTAAAATTTGGA 6393

Qy        636 CTTGCTGTATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCT 695

Db       6394 ATTGTTGTATATATTTTATATGTAAGATAGTATATAAGTAGAAATATTTAAAAGTAGTTT 6453

Qy        696 TAATTCAAGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTG 755

Db       6454 TATTTATAGATTGTAGTAATTTTGTATTTTTATTAAGATAATTTGTTTTGTGTTAAAATA 6513

Qy        756 TGGTTATTTTTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTT 815

Db       6514 GTAATTTTTAAATTTTTGTTTATTATGAAAAGGTAATTTTAAAGTTTATTATGTAAAATT 6573

Qy        816 ATGATGACAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGT 875

Db       6574 AATTATAAATAGGATTTAATTTATATTTATAGATTTTTTTAAGTATTATATAATTTAA-- 6631

Qy        876 GTATTTTTTATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTAT 935

Db       6632 AAATTTTTGTTTTAAAGTTTTGTTTTAATTATTGAAATATTTTAATTTGTGGTTATTAAT 6691

Qy        936 AGTGTTCTTTTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCA 995

Db       6692 TTAGTAAATTTAAGGAATTAGGTTATGATTAAG--------AATTTAGGTGGAATTGATG 6743

Qy        996 TTGCTATAATTCACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTAT 1055

Db       6744 TTTGGGTAATTAAAATAAATGGTATAAGAGTTTAAAAATTAAAGTTGTGTTAGTGGTTTT 6803

Qy       1056 GAAACAATATTAGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTT 1109

Db       6804 TAATTAGAGGTAGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTGTAGGGGT 6857
```

RESULT 6
AAS46788
ID AAS46788 standard; DNA; 61020 BP.
XX
AC AAS46788;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #514.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 514; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Sequences with even
CC numbered Seq ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 61020 BP; 17884 A; 634 C; 12537 G; 29965 T; 0 U; 0 Other;

Query Match 2.0%; Score 71.4; DB 4; Length 61020;
Best Local Similarity 48.2%; Pred. No. 4.6e-06;
Matches 201; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

```
Qy         517 TGGTGCTTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTA 576

Db       11783 TGTTTATTATAAATTTGTATATTTTATATTATTATTTTTAATTATATAATTAATATTTTTA 11842

Qy         577 AAAGGCTGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTAC 636

Db       11843 GGAAATTATATAGGTTTTTTAAGTTTAAAATATTTTTAATTAAATTTATTATTATTTTTT 11902

Qy         637 TTGCTGTATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTT 696

Db       11903 TAAATTTTTTTGTTTTTTAGTATTTTTTATTTAAGTAAATTATATAATTTTTTAAATTAT 11962

Qy         697 AATTCAAGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGT 756

Db       11963 GTTTGTGAGAAGTTTGGGATTTTTTTTTTTTTTTTTATTTTATTCGGTTATTGTTATTGT 12022

Qy         757 GGTTATTTTTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTA 816

Db       12023 TGTTATAATTATTTTTATATTTTGTTAATTTATTTTGTTTTTATATTTATTTATTTTTTT 12082

Qy         817 TGATGACAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTG 876

Db       12083 TGTTAGTTTTATTTTTTTTTTTTTTTTTAAATTTTAGAGTAATTTTATATTTTTTTTTT 12142

Qy         877 TATTTTTTATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTT 933

Db       12143 TGTTTTATATATTTTTTTTTTTTAAGTTTATTTTCGTTTGATATTTGTTATTTTATT 12199
```

RESULT 7
ADB54239
ID ADB54239 standard; DNA; 8666 BP.
XX
AC ADB54239;
XX
DT 04-DEC-2003 (first entry)
XX
DE Pretreated genomic DNA region 163.
XX
KW colon cell proliferative disorder; non methylated CpG dinucleotide;

KW cytostatic; cancer; adenoma; carcinoma; cytosine methylation state; ds.
XX
OS Unidentified.
XX
PN WO2003072821-A2.
XX
PD 04-SEP-2003.
XX
PF 27-FEB-2003; 2003WO-EP002035.
XX
PR 27-FEB-2002; 2002EP-00004551.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Adorjan P, Burger M, Maier S, Nimmrich I, Becker E, Lesche R;
PI Rujan T, Schmitt A;
XX
DR WPI; 2003-731620/69.
XX
PT Detecting and differentiating between colon cell proliferative disorders
PT associated with a gene or its regulatory regions comprises contacting a
PT target nucleic acid in a biological sample obtained from the subject with
PT a reagent.
XX
PS Claim 32; SEQ ID NO 295; 74pp; English.
XX
CC The invention relates to a novel method for detecting and differentiating
CC between colon cell proliferative disorders associated with at least one
CC gene or its regulatory regions. The method comprises contacting a target
CC nucleic acid in a biological sample obtained from the subject with at
CC least one reagent or a series of reagents, where the reagent or series of
CC reagents, distinguishes between methylated and non methylated CpG
CC dinucleotides within the target nucleic acid. The molecules of the
CC invention demonstrate cytostatic activity whilst the method may useful
CC for detecting and differentiating between colon cell proliferative
CC disorders, including cancers such as colon adenoma and colon carcinoma.
CC The PNA (peptide nucleic acid)-oligomers are useful as probes for
CC determining cytosine methylation state or single nucleotide
CC polymorphisms. The current sequence is that of the pretreated genomic DNA
CC region of the invention. This sequence is not shown within the
CC specification but is taken from Wipoweb.
XX
SQ Sequence 8666 BP; 2444 A; 0 C; 1868 G; 4354 T; 0 U; 0 Other;

Query Match 1.9%; Score 69.8; DB 9; Length 8666;
Best Local Similarity 45.0%; Pred. No. 5.7e-06;
Matches 263; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

```
Qy       525 TAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGCTG 584
Db        30 TAATTAGGTGTTTTTATTGTGTTATGATTTTTATTTTTAGTTGGGTTTTTTAGTTTGGTT 89

Qy       585 TTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTA 644
Db        90 TTAAAGAATAAAATAAAAGGTTTTTTTGTTTGAGTTTTATTTAATTTATTTTTTTTATTT 149

Qy       645 TTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAG 704
Db       150 ATAAGAGGTATTTTAAATTTTTATGATTTTAAATTTTTTAATTGATTATTTTATTTTTA 209

Qy       705 GTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTT 764
Db       210 TTTTGTAAATAAGTTAATTTTTTTTATTTATAGAAGGAAGTGTTTTAATTTTTTTTTTGTT 269

Qy       765 TTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACA 824
Db       270 TTGATTTTTTTTTTTTTTTTAAATTTATGTATTTTTTGTGATAAAATTTATAATTATTGTT 329

Qy       825 ACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTT 884
Db       330 GTATTTTGAGTTTTATTTTTTTATTATTTTTGAGGGATTTTAAGTTTTTAAAAATATTTT 389

Qy       885 ATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTT 944
Db       390 ATTTTGTTTGTGTATTTAATTTTTTTTTTTATTTTTTTTTAATTTTTTTTTTTTTTTATTT 449

Qy       945 TTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAA 1004
Db       450 GGTATTTGTTTTTTTAGGTATATGTGTGTTTAGGTTTTTTTTATTTTTTATTTGTTTTAT 509

Qy      1005 TTCACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATA 1064
Db       510 TTTATGGTATAGGGTTTTGAATTATTATAATTAAGTTATGAAAGAGTAGTTAATGTAGTG 569

Qy      1065 TTAGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTT 1109
Db       570 TTTTTATTTTTTTGTTATTTTATTATTTTAGTTTTTTTTTGGTT 614
```

RESULT 8
ADE84177
ID ADE84177 standard; DNA; 8666 BP.
XX
AC ADE84177;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human lymphoid cell proliferative disorder gene derived DNA #113.
XX
KW ds; lymphoid cell proliferative disorder; methylation;
KW methylated CpG dinucleotide; single nucleotide polymorphism; SNP;
KW diffuse large B-cell lymphoma; mantle cell lymphoma;
KW chronic lymphocytic leukemia; small lymphocytic lymphoma;
KW follicular lymphoma; diagnosis; prognosis.
XX
OS Homo sapiens.
XX
PN WO2003044226-A2.
XX
PD 30-MAY-2003.
XX
PF 25-NOV-2002; 2002WO-EP013265.
XX
PR 23-NOV-2001; 2001DE-01057491.
PR 28-DEC-2001; 2001DE-01064501.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Burger M, Caldwell C, Genc B, Becker E, Maier S, Nimmrich I;
XX
DR WPI; 2003-457621/43.
XX
PT Detecting and differentiating between lymphoid cell proliferative
PT disorders comprises contacting a target nucleic acid with at least one
PT reagent that distinguishes between methylated and non-methylated CpG .
PT dinucleotides.
XX
PS Claim 26; SEQ ID NO 173; 448pp; English.
XX
CC The invention relates to a method of detecting and differentiating
CC between lymphoid cell proliferative disorders associated with at least
CC one gene and/or their regulatory regions in a subject by contacting a
CC target nucleic acid in a biological sample obtained from the subject with
CC at least one reagent or series of reagents that distinguish between
CC methylated and non-methylated CpG dinucleotides within the target nucleic
CC acid. The genes and/or their regulatory regions are preferably selected
CC from MDR1, CSNK2B, EGR4, AR, CDK4, RB2, CDC25A, GPIb beta, MYOD1, CDH3,
CC MYCL1, ELK1, ABL1, APC, BCL2, CDH1, CDKN1A, CDKN1B, CDKN2a, CDKN2B, FOS,
CC GSTP1, HIC-1, MGMT, MLH1, MOS, MYC, PTEN, RBL2, TGFBR2, TP73, CDKNIC,
CC GSK3beta, ESR1, APAF1, BAK1, BAX or HOXA5. Oligomers, peptide nucleic
CC acid (PNA)-oligomers and/or isolated nucleic acids based on the sequences
CC of the genes are useful for detecting the methylation state of all the
CC CpG dinucleotides within one or more the sequences, or their complements,
CC for determining the cytosine methylation state and or single nucleotide
CC polymorphisms (SNPs), and for differentiating at least two of the medical
CC conditions such as diffuse large B-cell lymphoma, mantle cell lymphoma,

CC chronic lymphocytic leukemia, small lymphocytic lymphoma and follicular
CC lymphoma. They are also useful for detecting of a predisposition to,
CC differentiation between subclasses, diagnosis, prognosis, treating and/or
CC monitoring of lymphoid cell proliferative disorder. This sequence
CC represents a nucleic acid of a pretreated genomic DNA derived from the
CC above mentioned genes.
XX
SQ Sequence 8666 BP; 2444 A; 0 C; 1868 G; 4354 T; 0 U; 0 Other;

Query Match 1.9%; Score 69.8; DB 9; Length 8666;
Best Local Similarity 45.0%; Pred. No. 5.7e-06;
Matches 263; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

Qy 525 TAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGCTG 584
Db 30 TAATTAGGTGTTTTTATTGTGTTATGATTTTTATTTTTAGTTGGGTTTTTTAGTTTGGTT 89

Qy 585 TTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTA 644
Db 90 TTAAAGAATAAAATAAAAGGTTTTTTTGTTTGAGTTTTATTTAATTTATTTTTTTTATTT 149

Qy 645 TTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAG 704
Db 150 ATAAGAGGTATTTTAAATTTTTATGATTTTTAAATTTTTTAATTGATTATTTTATTTTTA 209

Qy 705 GTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTT 764
Db 210 TTTTGTAAATAAGTTAATTTTTTTATTTATAGAAGGAAGTGTTTTAATTTTTTTTTTGTT 269

Qy 765 TTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACA 824
Db 270 TTGATTTTTTTTTTTTTTTAAATTTATGTATTTTTTGTGATAAAATTTATAATTATTGTT 329

Qy 825 ACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTT 884
Db 330 GTATTTTGAGTTTTATTTTTTTATTATTTTTGAGGGATTTTAAGTTTTTAAAAATATTTT 389

Qy 885 ATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTT 944
Db 390 ATTTTGTTTGTGTATTTAATTTTTTTTTTATTTTTTTTTAATTTTTTTTTTTTTTTATTT 449

Qy 945 TTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAA 1004
Db 450 GGTATTTGTTTTTTTAGGTATATGTGTGTTTAGGTTTTTTTTATTTTTTATTTGTTTTAT 509

Qy 1005 TTCACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATA 1064
Db 510 TTTATGGTATAGGGTTTTGAATTATTATAATTAAGTTATGAAAGAGTAGTTAATGTAGTG 569

Qy 1065 TTAGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTT 1109
Db 570 TTTTTATTTTTTTGTTATTTTATTATTTTAGTTTTTTTTTTGGTT 614

RESULT 9
AAS46436
ID AAS46436 standard; DNA; 8093 BP.
XX
AC AAS46436;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #158.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 158; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Sequences with even
CC numbered Seq ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 8093 BP; 2419 A; 175 C; 1689 G; 3810 T; 0 U; 0 Other;

Query Match 1.9%; Score 69.4; DB 4; Length 8093;
Best Local Similarity 47.5%; Pred. No. 7e-06;
Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;

Qy 527 ATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGCTGTT 586
Db 6285 ATTGTTGTTTTATTTTAAAAAATTTTGTTTAAAGTAAATATTTTGTTTTAAAATGATATA 6344

Qy 587 TATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTATT 646
Db 6345 GTATTAGATTTTGTTAGATGTTAGAAATGGATTTATTTTAAAATTTGGAATTGTCGTATA 6404

Qy 647 GGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAGGT 706
Db 6405 TATTTTATATGTAAGATAGTATATAAGTAGAAATATTTAAAAGTAGTTTTATTTATAGAT 6464

Qy 707 AATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTTT 766
Db 6465 TGTAGTAATTTTGTATTTTTATTAAGATAATTTGTTTTGTGTTAAAATAGTAATTTTTAA 6524

Qy 767 ATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAAC 826
Db 6525 ATTTTTGTTTATTATGAAAAGGTAATTTTAAAGTTTATTATGTAAAATTAATTATAAATA 6584

Qy 827 ATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTTAT 886

```
Db       6585 GGATTTAATTTATATTTATAGATTTTTTTAAGTATTATATAATTT--AAAAATTTTTGTT 6642

Qy        887 ATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTT 946

Db       6643 TTAAAGTTTTGTTTTAATTATTGAAATATTTTAATTTGTGGTTATTAATTTAGTAAATTT 6702

Qy        947 AAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAATT 1006

Db       6703 AAGGAATTAGGTTATGATTAAG--------AATTTAGGTGGAATTGATGTTTGGGTAATT 6754

Qy       1007 CACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATATT 1066

Db       6755 AAAATAAATGGTATAAGAGTTTAAAAATTAAAGTTGTGTTAGTGGTTTTTAATTAGAGGT 6814

Qy       1067 AGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTT 1109

Db       6815 AGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTGTAGGGGT 6857
```

RESULT 10
ABK33974
ID ABK33974 standard; DNA; 8093 BP.
XX
AC ABK33974;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human DNA for staging of Astrocytomas, complement, #29.
XX
KW Human; ds; astrocytoma; cytostatic; staging; cysteine methylation; CpG;
KW bisulphite; brain tissue; MALDI; ESI; electron spray mass spectrometry;
KW matrix assisted laser desorption/ionization mass spectrometry.
XX
OS Homo sapiens.
XX
PN WO200202808-A2.
XX
PD 10-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007538.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-171649/22.
XX
PT Novel chemically modified genomic DNA sequences, useful in the
PT characterization, classification, differentiation, grading, staging,
PT treatment and/or diagnosis of astrocytomas or predisposition to
PT astrocytomas.
XX
PS Claim 1; SEQ ID NO 58; 37pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence (I) of at
CC least 18 bases in length of a segment of chemically pre-treated genomic
CC DNA which has any one of the sequences of (ABK33919-ABK34032) or its
CC complement. Also included are an oligonucleotide or peptide nucleic acid
CC (or set thereof) of at least 9 nucleotides which hybridises to (I),
CC primers for (I), probes for detecting cytosine methylation or single-
CC nucleotide polymorphisms (SNP) in (I), an array of oligomers or peptide
CC nucleic acids for analysing diseases associated with the methylation
CC states of the CpG dinucleotides of (I). The array is useful for
CC determining genetic and/or epigenetic parameters, classification,
CC differentiation, grading, staging, treatment and/or diagnosis of
CC astrocytomas, or the predisposition to astrocytomas by analysing cytosine
CC methylations, involves obtaining a biological sample containing genomic
CC DNA, extracting the genomic DNA, converting cytosine bases which are
CC unmethylated at the 5-position, in the genomic DNA sample, to uracil or
CC another base which is dissimilar to cytosine in terms of hybridisation
CC behaviour, by chemical treatment and amplifying chemically pre-treated
CC genomic DNA fragments using the array and a polymerase, where the
CC amplificates carry a detectable label. The method further involves
CC identifying methylation status of one or more cytosine positions, and
CC analysing methylation status of the cytosine positions by reference to
CC one or more data sets. The genomic DNA is chemically treated by using a
CC bisulphite, hydrogen sulphite or disulphite. The amplification step
CC amplifies DNA which is of particular interest in astrocytoma or brain
CC tissue, based on the specific genomic methylation status of brain
CC tissues, as opposed to background DNA. The amplificates carry a
CC fluorescent label or radionuclide. Optionally, the labels of the
CC amplificates are detachable molecule fragments having a typical mass
CC which are detected in a mass spectrometer. The fragments of chemically
CC pre-treated genomic DNA to be amplified, have a single positive or
CC negative charge for a better detectability in the mass spectrometer.
CC Preferably, the amplificates or fragments of the amplificates are
CC detected by matrix assisted laser desorption/ionization mass spectrometry
CC (MALDI) or using electron spray mass spectrometry (ESI). The present
CC sequence is one of the chemically pre-treated reference DNA samples of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 8093 BP; 2419 A; 175 C; 1689 G; 3810 T; 0 U; 0 Other;

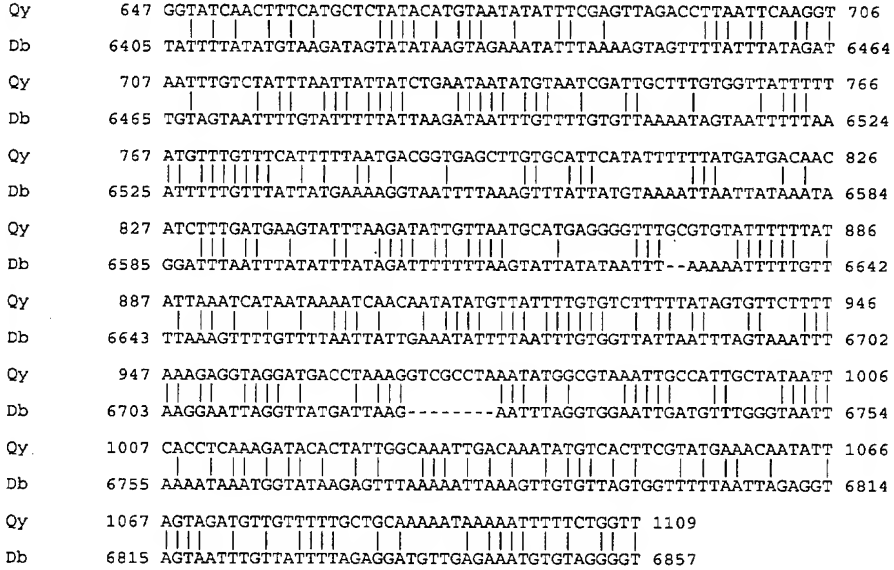
Query Match 1.9%; Score 69.4; DB 6; Length 8093;
Best Local Similarity 47.5%; Pred. No. 7e-06;
Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;

```
Qy        527 ATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGCTGTT 586

Db       6285 ATTGTTGTTTTATTTTAAAAAATTTGTTTAAAGTAAATATTTTGTTTTAAAATGATATA 6344

Qy        587 TATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTATT 646

Db       6345 GTATTAGATTTTGTTAGATGTTAGAAATGGATTTATTTTAAAATTTGGAATTGTCGTATA 6404

Qy        647 GGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAGGT 706

Db       6405 TATTTTATATGTAAGATAGTATATAAGTAGAAATATTTAAAAGTAGTTTTATTTATAGAT 6464

Qy        707 AATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTTT 766

Db       6465 TGTAGTAATTTTGTATTTTTATTAAGATAATTTGTTTTGTGTTAAAATAGTAATTTTTAA 6524

Qy        767 ATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAAC 826

Db       6525 ATTTTTGTTTATTATGAAAAGGTAATTTTAAAGTTTATTATGTAAAATTAATTATAAATA 6584

Qy        827 ATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTTAT 886

Db       6585 GGATTTAATTTATATTTATAGATTTTTTTAAGTATTATATAATTT--AAAAATTTTTGTT 6642

Qy        887 ATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTT 946

Db       6643 TTAAAGTTTTGTTTTAATTATTGAAATATTTTAATTTGTGGTTATTAATTTAGTAAATTT 6702

Qy        947 AAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAATT 1006

Db       6703 AAGGAATTAGGTTATGATTAAG--------AATTTAGGTGGAATTGATGTTTGGGTAATT 6754

Qy       1007 CACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATATT 1066

Db       6755 AAAATAAATGGTATAAGAGTTTAAAAATTAAAGTTGTGTTAGTGGTTTTTAATTAGAGGT 6814

Qy       1067 AGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTT 1109

Db       6815 AGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTGTAGGGGT 6857
```

RESULT 11
ABL92237

ID ABL92237 standard; DNA; 8093 BP.
XX
AC ABL92237;
XX
DT 01-JUL-2002 (first entry)
XX
DE Chemically treated DNA repair gene fragment complementary to#23.
XX
KW DNA repair; cytosine methylation; PMS2L1; PMS2L12; PMS2L2; PMS2L3; PMS2;
KW L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; TDG; INPPL1; RFC4; DDIT1L; FANCB;
KW XRCC8; ataxia telangiectasia; aging; Bloom's syndrome; Cockayne syndrome;
KW Nijmegen breakage syndrome; Werner syndrome; immunodeficiency;
KW trichthiodystrophy; Fanconi's anaemia; solid tumour; cancer; ds.
XX
OS Unidentified.
XX
PN WO200181622-A2.
XX
PD 01-NOV-2001.
XX
PF 06-APR-2001; 2001WO-EP003972.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-034446/04.
XX
PT New nucleic acid derived from genes associated with DNA repair, useful
PT for diagnosis, e.g. of ataxia telangiectasia, by determination of
PT cytosine methylation.
XX
PS Claim 1; SEQ ID NO 46; 25pp + Sequence Listing; English.
XX
CC The invention relates to nucleic acids containing a sequence of at least
CC 18 nucleotides of chemically treated DNA of genes associated with DNA
CC repair, and their complements. The invention also relates to nucleic
CC acids comprising at least 18 base pairs of the chemically pretreated DNA
CC of genes associated with DNA repair selected from PMS2L1, PMS2L12,
CC PMS2L2, PMS2L3, PMS2, L4, PMS2L5, PMS2L6, MGMT, MSH2, NUDT1, TDG, INPPL1,
CC RFC4, DDIT1L, FANCB, or XRCC8. Nucleic acids of the invention and related
CC oligomers, are useful for diagnosis of diseases associated with gene
CC repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,
CC Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,
CC immunodeficiency, trichthiodystrophy, Fanconi's anaemia, solid tumours
CC and cancer, particularly by determining status of cytosine methylation
CC and/or by detecting single-nucleotide polymorphisms. Determination of
CC individual methylation patterns may allow development of individualised
CC therapies. The sequences given in records ABL92192-ABL92335 represent
CC chemically pre-treated DNA fragments from genes associated with DNA
CC repair, and their complements. Note: The sequence data for this patent is
CC not represented in the specification, but is based on sequence
CC information supplied by the European Patent Office
XX
SQ Sequence 8093 BP; 2419 A; 175 C; 1689 G; 3810 T; 0 U; 0 Other;

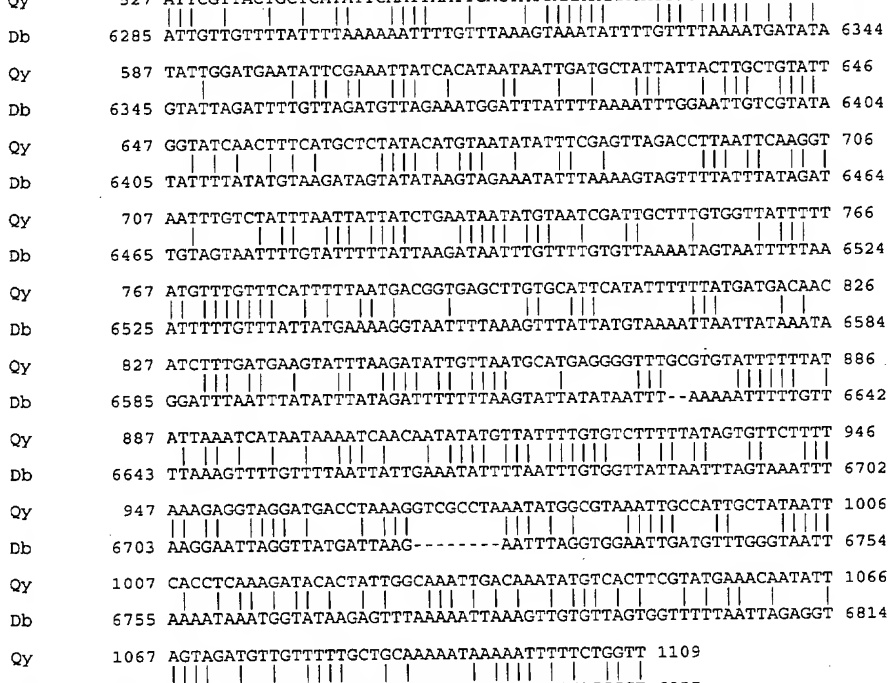
Query Match 1.9%; Score 69.4; DB 6; Length 8093;
Best Local Similarity 47.5%; Pred. No. 7e-06;
Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;



RESULT 12
ABL49332
ID ABL49332 standard; DNA; 8093 BP.
XX
AC ABL49332;
XX
DT 01-MAY-2002 (first entry)
XX
DE Human polynucleotide associated with DNA replication SEQ ID NO 32.
XX
KW Human; cytostatic; neuroprotective; nootropic; immunostimulant;
KW gene therapy; gene regulation; DNA replication; CENPB; DNA2L; ATR; CHD1L;
KW ERCC3; SNRPA1; RAD50; LIG2; cytosine methylation; Ataxia telangiectasia;
KW ATR-X; Bloom's syndrome; tumour; cancer; methylation; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200177377-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP003971.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017471/02.
XX
PT New nucleic acid sequences from chemically modified genes associated with
PT DNA replication, useful for analyzing cytosine methylations for diagnosis
PT and therapy of diseases e.g. Ataxia telangiectasia.
XX
PS Claim 1; SEQ ID NO 32; 23pp; English.
XX

CC The invention relates to nucleic acid sequences comprising at least 18
CC bases of a chemically pretreated gene associated with gene regulation,
CC selected from 94 genes (ABL49301-ABL49394) and/or complementary sequences
CC associated with DNA replication, CENPB, DNA2L, ATR, CHD1L, ERCC3, SNRPA1,
CC RAD50 and LIG2. The chemical pretreatment converts cytosine bases
CC unmethylated at the 5-position to uracil or another base with
CC hybridisation behaviour dissimilar to cytosine, to enable analysis of
CC cytosine methylations. The DNA sequences and method are useful in the
CC diagnosis of diseases (or predisposition to diseases) associated with DNA
CC replication and in therapy of such diseases, by enabling analysis of the
CC cytosine methylation patterns of such genes. They are especially useful
CC in diagnosis and therapy of e.g. Ataxia telangiectasia, ATR-X, Bloom's
CC syndrome, solid tumours and cancer. Note: The sequence data for this
CC patent did not form part of the printed specification and was supposed to
CC be available directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences. However, the sequence data did
CC not correspond to that referred to in the specification. The present data
CC is taken from EPO data for the patent
XX
SQ Sequence 8093 BP; 2419 A; 175 C; 1689 G; 3810 T; 0 U; 0 Other;

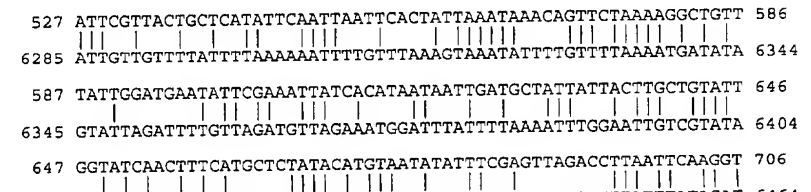
Query Match 1.9%; Score 69.4; DB 6; Length 8093;
Best Local Similarity 47.5%; Pred. No. 7e-06;
Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;



RESULT 13
ABZ10032
ID ABZ10032 standard; DNA; 8093 BP.
XX
AC ABZ10032;
XX
DT 16-JAN-2003 (first entry)

XX
DE Haematopoietic cell proliferation disorder related DNA sequence #172.
XX
KW Human; haematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW cytosine methylation state; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277272-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-EP003401.
XX
PR 26-MAR-2001; 2001US-0278333P.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI Schwope I, Ziebarth H;
XX
DR WPI; 2003-018942/01.
XX
PT Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent that
PT distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS Claim 28; SEQ ID NO 172; 117pp; English.
XX
CC The present invention describes a method for detecting and
CC differentiating between haematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used: for
CC differentiating between healthy haematopoietic cells and proliferative
CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related DNA
CC sequences. The nucleotide sequences from the present invention can also
CC be used for detecting a predisposition to, differentiation between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC haematopoietic cell proliferative disorders. The present method enables a
CC highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients
XX
SQ Sequence 8093 BP; 2419 A; 175 C; 1689 G; 3810 T; 0 U; 0 Other;

Query Match 1.9%; Score 69.4; DB 7; Length 8093;
Best Local Similarity 47.5%; Pred. No. 7e-06;
Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;



```
Qy     707 AATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTTT 766
Db    6465 TGTAGTAATTTTGTATTTTTATTAAGATAATTTGTTTTGTGTTAAAATAGTAATTTTTAA 6524

Qy     767 ATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAAC 826
Db    6525 ATTTTTGTTTATTATGAAAAGGTAATTTTAAAGTTTATTATGTAAAATTAATTATAAATA 6584

Qy     827 ATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTTAT 886
Db    6585 GGATTTAATTTATATTTATAGATTTTTTTAAGTATTATATAATTT--AAAAATTTTTGTT 6642

Qy     887 ATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTT 946
Db    6643 TTAAAGTTTTGTTTTAATTATTGAAATATTTTAATTTGTGGTTATTAATTTAGTAAATTT 6702

Qy     947 AAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAATT 1006
Db    6703 AAGGAATTAGGTTATGATTAAG--------AATTTAGGTGGAATTGATGTTTGGGTAATT 6754

Qy    1007 CACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATATT 1066
Db    6755 AAAATAAATGGTATAAGAGTTTAAAAAATTAAAGTTGTGTTAGTGGTTTTTAATTAGAGGT 6814

Qy    1067 AGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTT 1109
Db    6815 AGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTGTAGGGGT 6857
```

RESULT 14
ADA20361
ID ADA20361 standard; DNA; 8093 BP.
XX
AC ADA20361;
XX
DT 20-NOV-2003 (first entry)
XX
DE Prostate tumour related genomic DNA complement sample #13.
XX
KW cytostatic; gene therapy; genetic marker; epigenetic parameter;
KW classification; differentiation; diagnosis; prostate tumour;
KW prostate cancer; cytosine methylation; uracil;
KW single nucleotide polymorphism; SNP; prostate carcinoma; ss.
XX
OS Homo sapiens.
XX
PN WO2002103042-A2.
XX
PD 27-DEC-2002.
XX
PF 14-JUN-2002; 2002WO-EP006605.
XX
PR 14-JUN-2001; 2001DE-01028508.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Distler J, Model F, Adorjan P;
XX
DR WPI; 2003-167536/16.
XX
PT Determining genetic and/or epigenetic parameters, useful for the
PT classification, differentiation and/or diagnosis of prostate tumors or a
PT predisposition to prostate cancer, comprises analyzing cytosine
PT methylation.
XX
PS Claim 28; Page 128-132; 376pp; English.
XX
CC The invention relates to a method of determining genetic and/or
CC epigenetic parameters for the classification, differentiation and/or
CC diagnosis of prostate tumours or the predisposition to prostate cancer,
CC by analysing cytosine methylation in a sample of genomic DNA. The method
CC comprises chemically treating unmethylated cytosine bases at the 5-
CC position to uracil or another base, which is dissimilar to cytosine in
CC terms of hybridization behaviour; followed by amplifying at least one
CC fragment of the chemically pre-treated genomic DNA using sets of primer
CC oligonucleotides and a polymerase. The oligomers or probes derived from
CC them are useful for detecting the methylation state of all CpG
CC dinucleotides and/or single nucleotide polymorphisms (SNPs) in a
CC chemically pre-treated genomic DNA. They are all useful for treating
CC prostate carcinoma. This sequence represents a fragment of genomic DNA
CC used in the method of the invention.
XX
SQ Sequence 8093 BP; 2419 A; 175 C; 1689 G; 3810 T; 0 U; 0 Other;

Query Match 1.9%; Score 69.4; DB 7; Length 8093;
Best Local Similarity 47.5%; Pred. No. 7e-06;
Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;

```
Qy     527 ATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGCTGTT 586
Db    6285 ATTGTTGTTTATTTTAAAAAATTTTGTTTAAAGTAAATATTTTGTTTTAAAATGATATA 6344

Qy     587 TATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTATT 646
Db    6345 GTATTAGATTTTGTTAGATGTTAGAAATGGATTTATTTTAAAATTTGGAATTGTCGTATA 6404

Qy     647 GGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAGGT 706
Db    6405 TATTTTATATGTAAGATAGTATATAAGTAGAAATATTTAAAAGTAGTTTTATTTATAGAT 6464

Qy     707 AATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTTT 766
Db    6465 TGTAGTAATTTTGTATTTTTATTAAGATAATTTGTTTTGTGTTAAAATAGTAATTTTTAA 6524

Qy     767 ATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAAC 826
Db    6525 ATTTTTGTTTATTATGAAAAGGTAATTTTAAAGTTTATTATGTAAAATTAATTATAAATA 6584

Qy     827 ATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTTAT 886
Db    6585 GGATTTAATTTATATTTATAGATTTTTTTAAGTATTATATAATTT--AAAAATTTTTGTT 6642

Qy     887 ATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTT 946
Db    6643 TTAAAGTTTTGTTTTAATTATTGAAATATTTTAATTTGTGGTTATTAATTTAGTAAATTT 6702

Qy     947 AAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAATT 1006
Db    6703 AAGGAATTAGGTTATGATTAAG--------AATTTAGGTGGAATTGATGTTTGGGTAATT 6754

Qy    1007 CACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATATT 1066
Db    6755 AAAATAAATGGTATAAGAGTTTAAAAAATTAAAGTTGTGTTAGTGGTTTTTAATTAGAGGT 6814

Qy    1067 AGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTT 1109
Db    6815 AGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTGTAGGGGT 6857
```

RESULT 15
ADA84168
ID ADA84168 standard; DNA; 8093 BP.
XX
AC ADA84168;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human renal/prostate carcinoma associated DNA SEQ ID NO:26.
XX
KW ds; renal cancer; prostate cancer; cytosine methylation;
KW single nucleotide polymorphism; histological; cytological.
XX
OS Homo sapiens.
XX
PN WO2002103041-A2.
XX

PD 27-DEC-2002.
XX
PF 14-JUN-2002; 2002WO-EP006603.
XX
PR 14-JUN-2001; 2001DE-01028509.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Distler J, Model F, Adorjan P;
XX
DR WPI; 2003-183991/18.
XX
PT Method for characterizing, classifying and/or differentiating renal and
PT prostate cancers, by analyzing the genetic and/or epigenetic parameters
PT of genomic DNA, particularly by determining its cytosine methylation
PT status.
XX
PS Claim 1; Page 80-83; 211pp; English.
XX
CC The invention relates to a novel method for characterising, classifying
CC and/or differentiating renal and prostate cancer. The method comprises
CC extracting genomic DNA from a biological sample, converting cytosine
CC bases (by chemical treatment) that are unmethylated at the 5-position to
CC uracil or another base, and amplifying at least one fragment of the
CC chemically pretreated genomic DNA using sets of primer oligonucleotides
CC and a polymerase. The method is useful for detecting the cytosine
CC methylation state and/or single nucleotide polymorphisms in genomic DNA,
CC particularly for characterising, classifying and/or differentiating renal
CC and prostate cancers. The oligomers are useful as primer oligonucleotides
CC for the amplification of any of the 112 DNA sequences of the invention.
CC The set of oligomer probes is useful for detecting the cytosine
CC methylation state and/or single nucleotide polymorphisms in any of the
CC 112 chemically pretreated genomic DNA sequences. The method is also
CC useful for identifying the tissue of origin of cancer cells. The method
CC allows the classification, differentiation and/or diagnosis of cancer
CC tissues using minute samples which would be inadequate for histological
CC or cytological analysis. The present sequence represents one of the 112
CC DNA sequences of the invention.
XX
SQ Sequence 8093 BP; 2419 A; 175 C; 1689 G; 3810 T; 0 U; 0 Other;

Query Match 1.9%; Score 69.4; DB 7; Length 8093;
Best Local Similarity 47.5%; Pred. No. 7e-06;
Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;

```
Qy        527 ATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGCTGTT 586
              |||  |  |  |  ||   ||||   |    | ||||||    ||| |||| | | |
Db       6285 ATTGTTGTTTTATTTTAAAAAATTTTGTTTAAAGTAAATATTTTGTTTTAAAATGATATA 6344

Qy        587 TATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTATT 646
              |      | || ||  |||  |   ||   |  |   |||   | |||  ||||
Db       6345 GTATTAGATTTTGTTAGATGTTAGAAATGGATTTATTTTAAAATTTGGAATTGTCGTATA 6404

Qy        647 GGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAGGT 706
              | | |  | | |    ||||  | |||  |   ||  |          ||| ||  || |
Db       6405 TATTTTATATGTAAGATAGTATATAAGTAGAAATATTTAAAAGTAGTTTTATTTATAGAT 6464

Qy        707 AATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTTT 766
              |      | ||  |||  ||||    |||||  |||  |   ||    |     | |||
Db       6465 TGTAGTAATTTTGTATTTTTATTAAGATAATTTGTTTTGTGTTAAAATAGTAATTTTTAA 6524

Qy        767 ATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAAC 826
              || |||||||  | |  || |     |       ||  |||         |||     | |
Db       6525 ATTTTTGTTTATTATGAAAAGGTAATTTTAAAGTTTATTATGTAAAATTAATTATAAATA 6584

Qy        827 ATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTTAT 886
              |||  ||  |   ||  |||| || ||||    |       |||       ||||||  |
Db       6585 GGATTTAATTTATATTTATAGATTTTTTTAAGTATTATATAATTT--AAAAATTTTTGTT 6642

Qy        887 ATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTT 946
              | || |  |  |  ||| |  |.|||| ||| ||||||  | || ||   ||   |||
Db       6643 TTAAAGTTTTGTTTTAATTATTGAAATATTTTAATTTGTGGTTATTAATTTAGTAAATTT 6702

Qy        947 AAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAATT 1006
              || ||  |||| |    | |||         ||| | |    |||||  ||   |||||
Db       6703 AAGGAATTAGGTTATGATTAAG--------AATTTAGGTGGAATTGATGTTTGGGTAATT 6754

Qy       1007 CACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATATT 1066
              | | || | || |  | |   ||| | | |  | ||| | |   | | ||  |     |
Db       6755 AAAATAAATGGTATAAGAGTTTAAAAATTAAAGTTGTGTTAGTGGTTTTTAATTAGAGGT 6814

Qy       1067 AGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTT 1109
              ||||  | |  ||||   | |      | |||| | |  || |
Db       6815 AGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTGTAGGGGT 6857
```

Search completed: July 2, 2004, 16:42:07
Job time : 1317 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 2, 2004, 15:49:42 ; Search time 237 Seconds
(without alignments)
8401.535 Million cell updates/sec

Title: US-10-780-347-1
Perfect score: 3588
Sequence: 1 gtcgacttattgcattgatg.....gatnaaattgagccgcttaa 3588

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	69.4	1.9	8093	4	US-10-204-708-32	Sequence 32, Appl
	2	68.6	1.9	837	3	US-08-998-416-288	Sequence 288, App
c	3	66	1.8	19124	2	US-08-487-826B-13	Sequence 13, Appl
	4	65.2	1.8	6040	4	US-10-204-708-69	Sequence 69, Appl
	5	65	1.8	658	3	US-08-998-416-595	Sequence 595, App
	6	63	1.8	717	4	US-09-252-991A-8972	Sequence 8972, Ap
	7	63	1.8	819	4	US-09-252-991A-8861	Sequence 8861, Ap
	8	60.4	1.7	5562	4	US-10-204-708-63	Sequence 63, Appl
	9	60	1.7	6866	4	US-10-204-708-20	Sequence 20, Appl
	10	59.8	1.7	640681	4	US-09-790-988-1	Sequence 1, Appli
	11	59.4	1.7	600	4	US-09-252-991A-752	Sequence 752, App
	12	59.4	1.7	681	4	US-09-252-991A-723	Sequence 723, App
c	13	59.4	1.7	2019	4	US-09-252-991A-658	Sequence 658, App
	14	59.2	1.6	665	2	US-08-883-795A-36	Sequence 36, Appl
	15	57.6	1.6	11015	4	US-10-204-708-55	Sequence 55, Appl
c	16	56	1.6	579	4	US-09-252-991A-9199	Sequence 9199, Ap
c	17	55.6	1.5	8920	2	US-08-446-855A-1	Sequence 1, Appli
c	18	55.6	1.5	8920	3	US-09-150-741-1	Sequence 1, Appli
	19	55.2	1.5	6113	4	US-10-204-708-14	Sequence 14, Appl
	20	55	1.5	10467	4	US-10-204-708-2	Sequence 2, Appli
	21	54.6	1.5	53332	4	US-09-801-861-3	Sequence 3, Appli
c	22	53	1.5	832	4	US-09-621-976-2813	Sequence 2813, Ap
	23	53	1.5	6669	4	US-10-204-708-6	Sequence 6, Appli
	24	53	1.5	19124	2	US-08-487-826B-13	Sequence 13, Appl
c	25	52.8	1.5	6243	2	US-09-056-075-1	Sequence 1, Appli
	26	52.6	1.5	5152	4	US-10-204-708-73	Sequence 73, Appl
	27	52.6	1.5	5501	4	US-10-204-708-38	Sequence 38, Appl
	28	52.4	1.5	7304	4	US-10-204-708-43	Sequence 43, Appl
	29	52	1.4	6306	4	US-10-204-708-50	Sequence 50, Appl
	30	52	1.4	6326	4	US-10-204-708-57	Sequence 57, Appl
	31	51.8	1.4	8961	4	US-10-204-708-80	Sequence 80, Appl
	32	51.4	1.4	396	4	US-09-640-173-53	Sequence 53, Appl
	33	51.4	1.4	396	4	US-09-713-550-53	Sequence 53, Appl
	34	51.4	1.4	19233	4	US-10-204-708-45	Sequence 45, Appl
	35	51.2	1.4	11131	4	US-10-204-708-28	Sequence 28, Appl
	36	50.8	1.4	8607	4	US-10-204-708-72	Sequence 72, Appl
c	37	50.6	1.4	2347	4	US-08-956-171E-153	Sequence 153, App
	38	50	1.4	8607	4	US-10-204-708-71	Sequence 71, Appl
	39	49.8	1.4	615	3	US-08-998-416-186	Sequence 186, App
	40	49.4	1.4	354	4	US-09-328-352-2580	Sequence 2580, Ap
	41	49.4	1.4	9347	4	US-10-204-708-35	Sequence 35, Appl
c	42	49.4	1.4	640681	4	US-09-790-988-1	Sequence 1, Appli
	43	49.2	1.4	6156	4	US-10-204-708-60	Sequence 60, Appl
	44	49	1.4	5666	4	US-10-204-708-29	Sequence 29, Appl
	45	49	1.4	19233	4	US-10-204-708-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1
US-10-204-708-32
; Sequence 32, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 32
; LENGTH: 8093
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-32

Query Match 1.9%; Score 69.4; DB 4; Length 8093;
Best Local Similarity 47.5%; Pred. No. 4.8e-08;
Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;

```
Qy       527 ATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGCTGTT 586
             ||| |  |  | ||  |||| |  | ||||||   ||| ||||| | | |
Db      6285 ATTGTTGTTTTATTTTAAAAAATTTTGTTTAAAGTAAATATTTTGTTTTAAAATGATATA 6344

Qy       587 TATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTATT 646
             |       | || || ||| |   ||  |  |  ||| | |||  ||||
Db      6345 GTATTAGATTTTGTTAGATGTTAGAAATGGATTTATTTTAAAATTTGGAATTGTCGTATA 6404

Qy       647 GGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAGGT 706
               | | |  | | |     |||| | ||| |  ||  |       ||| ||  || |
Db      6405 TATTTTATATGTAAGATAGTATATAAGTAGAAATATTTAAAAGTAGTTTTATTTATAGAT 6464

Qy       707 AATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTTT 766
               |    | ||  ||| ||||   ||||| |||  |  ||   |    | |||
```

```
Db       6465 TGTAGTAATTTTGTATTTTTATTAAGATAATTTGTTTTGTGTTAAAATAGTAATTTTTAA 6524

Qy        767 ATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAAC 826

Db       6525 ATTTTTGTTTATTATGAAAAGGTAATTTTAAAGTTTATTATGTAAAATTAATTATAAATA 6584

Qy        827 ATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTTAT 886

Db       6585 GGATTTAATTTATATTTATAGATTTTTTTAAGTATTATATAATTT--AAAAATTTTTGTT 6642

Qy        887 ATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTT 946

Db       6643 TTAAAGTTTTGTTTTAATTATTGAAATATTTTAATTTGTGGTTATTAATTTAGTAAATTT 6702

Qy        947 AAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAATT 1006

Db       6703 AAGGAATTAGGTTATGATTAAG--------AATTTAGGTGGAATTGATGTTTGGGTAATT 6754

Qy       1007 CACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATATT 1066

Db       6755 AAAATAAATGGTATAAGAGTTTAAAAATTAAAGTTGTGTTAGTGGTTTTTAATTAGAGGT 6814

Qy       1067 AGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTT 1109

Db       6815 AGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTGTAGGGGT 6857
```

RESULT 2
US-08-998-416-288
; Sequence 288, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 288:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1241RP
US-08-998-416-288

Query Match 1.9%; Score 68.6; DB 3; Length 837;
Best Local Similarity 47.6%; Pred. No. 2.7e-08;
Matches 265; Conservative 0; Mismatches 289; Indels 3; Gaps 2;

```
Qy        524 TTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGCT 583

Db        154 TTAATTCATCATTATTAATATTTATATAATTATAAAAATAATATTTAATATGAATACTAT 213

Qy        584 GTTTATTGGATGAATATTCGAAATTA--TCACATAATAATTGATGCTATTATTACTTGCT 641

Db        214 TTAGTCTATGTTCAAATTTTAAATTAGTTATTAAAATATTATTAGATATTATTATTTTCT 273

Qy        642 GTATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTC 701

Db        274 TTAAT-AAATTATTAAATAGATTATCAATAATTAATATATTATTTATTAATTGTTTATTA 332

Qy        702 AAGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTA 761

Db        333 AAATAATATATTTTATTATTATAAAGATTTAATTTATTTAAATATTGTAAATTATTATTT 392

Qy        762 TTTTTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATG 821

Db        393 TTATTATAATATCTATTTTTATAAATATTATGTTGATTTATATTATTTAATCTTTTTATA 452

Qy        822 ACAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTT 881

Db        453 AGAATTATTATTAAAATTAATTTTAACTTTAATTTCTTATTATTAATTTTTATATTATTT 512

Qy        882 TTTATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTT 941

Db        513 AATAAATTATATTTCATTTTATTTATTTATTTATTTAATTAAATTAATTATTTAATTAAT 572

Qy        942 CTTTTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTA 1001

Db        573 ATTTTATCATTATTTAATTAATTAATAAAATATTATAAAGAATGTAGTTAAAAATACTTA 632

Qy       1002 TAATTCACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACA 1061

Db        633 TAAAAGGATCCGAACCTATATTATTGTTTATGAGACAAATGCTTTAGCCCATAAGCTATA 692

Qy       1062 ATATTAGTAGATGTTGT 1078

Db        693 TAGTTTGACTATCATTT 709
```

RESULT 3
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellems, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,826B
;      FILING DATE:  10-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-487-826B-13

  Query Match             1.8%;  Score 66;  DB 2;  Length 19124;
  Best Local Similarity   47.6%;  Pred. No. 5.7e-07;
  Matches  195;  Conservative    0;  Mismatches  215;  Indels    0;  Gaps    0;

Qy        523 TTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGC 582
              |||||||  ||  |  | |  |  | ||  ||   |||||  |       | |||
Db      15903 TTTAATTTTTTTTTATTTAATTAAATTTTTTTATTTATTATTTTTTTTTATTTAATTAAA 15844

Qy        583 TGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTG 642
              | ||| |      || ||    | | |   |||  ||  |  | |   ||  |  |
Db      15843 TTTTTTTATTTATTATTTTTTTTTTATTTAAATAAATTTTTTTTATTTTATGTATATATTT 15784

Qy        643 TATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCA 702
              | |||  || ||  |||      | | |      |     || |  | ||  |  |||||  |
Db      15783 TTTTTTTAACATTTTTTTAATTTTTTTTTATTTTATGATATATATTTTTATTTTAATATA 15724

Qy        703 AGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTAT 762
                |  | | | | |||  || ||||    |   || | |  |   ||  | | |  || |
Db      15723 TTTTTTTCTTTTTTTTTTGTTTTTATGATATATATTTTTTTTTTTTTTAATGTTTTTTTTT 15664

Qy        763 TTTTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGA 822
              || |   |||||||    |||| | |   | | ||  |   |||   ||| | | |
Db      15663 TTCTTCTTTTGTTTTTATTTTTTTTTATAATCATTTTTTTTTTTATATAAAATTTTTTTTTA 15604

Qy        823 CAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTT 882
              | |||||| |  | |||  | | || ||    ||  |   | |     |||  |
Db      15603 ATTTTTTTTTGATAATCTTTTTCATTTTTTATTCTATCAAAATTTATATTTTATTATAAT 15544

Qy        883 TTATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTT 932
              || |||||  |  |||  |||  |   |    | | || |||| | | ||||
Db      15543 TTTTATTATTTTTTAAAAAATTTTTCTCCTTTTTTTTTTTTTTTTTTTT 15494


RESULT 4
US-10-204-708-69
; Sequence 69, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 69
;   LENGTH: 6040
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-69

  Query Match             1.8%;  Score 65.2;  DB 4;  Length 6040;
  Best Local Similarity   46.0%;  Pred. No. 5.3e-07;
  Matches  261;  Conservative    0;  Mismatches  303;  Indels    4;  Gaps    1;

Qy        548 ATTAATTCACTATTAAATAAACAGTTCTAAAAGGCTGTTTATTGGATGAATATTCGAAAT 607
              |||| ||   | || ||  |   | | |   | | ||  ||  | |  | |  || |
Db       1395 ATTAGTTGGTTTTTTAAAATGTATATTTTTTATTTTTTTGTTTTAAAGTTTTTATTAAGT 1454

Qy        608 TATCACATAATAATTGATGCTATTATTACTTGCTGTATTGGTATCAACTTTCATGCTCTA 667
              | | |||  |||  |   ||    | ||  ||   ||||||  || |||  |    |
Db       1455 AAGTAATTAAATATTATTAAAATATAAAGTTTTTGGTATGGTATAAAGTTTTTTAAAATT 1514

Qy        668 TACATGTAATATATTTCGAGTTAGACCTTAATTCAAGGTAATTT----GTCTATTTAATT 723
              |   | |||| | ||| |  |||     ||| |       | ||  |     | | ||| |||
Db       1515 TGGTTTTAATTTTTTTAGTTTTATTTATTATTAGTTTATTATATTTGATTTTTTTTTATT 1574

Qy        724 ATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTTTATGTTTGTTTCATTTTT 783
              || |   | |||  | | ||  | |  ||     |||| ||| |  || ||  | || |
Db       1575 TTTGTTATATTAAGGTTTTATGTAATTATTGAATTTTATGTTTTTTGTTTTTATAATTGT 1634

Qy        784 AATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAACATCTTTGATGAAGTATT 843
              ||      | |  |    | ||  | ||| ||  |             |  ||||    | ||
Db       1635 TATTTTTTTTATTTACGGGATAGAGATTGTTGTTTTATTTTAGTTAATGATTTTTTTTT 1694

Qy        844 TAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTTATATTAAATCATAATAAAA 903
              ||| || || | ||      | | |   | |  |||||  |      ||  |  | ||
Db       1695 TAATATTTTTTAAAGTATAGGTGAGAGAGAATTATTTTTTTTTTGGGGATATTTTTTAAG 1754

Qy        904 TCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTTAAAGAGGTAGGATGACC 963
              |     ||  |  |||| | | | || || |  |     ||    |  ||  |
Db       1755 TAGTTTTTAGTTAGTATTATTTTTTTTTATTGTGTTTAATGATTTTTTATATATTTTTTGT 1814

Qy        964 TAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAATTCACCTCAAAGATACACT 1023
              || | |   |    || ||   || |||   ||      ||| |  | ||  ||  | |
Db       1815 TATAAGAGGAGAGAAAATATTTTATATTTATTTTTTGGAGATTTAGTTTAATTATGAAAT 1874

Qy       1024 ATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATATTAGTAGATGTTGTTTTTG 1083
              | || || |      | |   |     | |||||     |||||  ||   |     |
Db       1875 GTGGGTAATTATTTTTGTTTTGTAGAGTTTGTGAAAATTATTTAGTGTATTAAGAAAAAG 1934

Qy       1084 CTGCAAAAATAAAAATTTTTCTGGTTGA 1111
              |   ||||||||  |||||   ||| |
Db       1935 ATATGAAAATAAATTTTTTTAGAGTTAA 1962


RESULT 5
US-08-998-416-595
; Sequence 595, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
```

```
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  595:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  658 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1408RP
US-08-998-416-595

  Query Match             1.8%;  Score 65;  DB 3;  Length 658;
  Best Local Similarity   49.0%;  Pred. No. 2.1e-07;
  Matches  173;  Conservative    0;  Mismatches  180;  Indels    0;  Gaps    0;

Qy        594 TGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTATTGGTATCA 653
Db        210 TTATTATTTTGTTATTTTATTATTTTATTATTTTATTATTTTATTATTTTATTATTTTAT 269

Qy        654 ACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAGGTAATTTGT 713
Db        270 TATTTTATTATTTTATTATTTATTTTATTATTTATTTTATTATTTATTTTATTATTTATT 329

Qy        714 CTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTTTATGTTTG 773
Db        330 TTATTATTTATTTATTTATTATTTATTTATTTGTTTGTTTGTTTATTATTTTTTTATTTA 389

Qy        774 TTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAACATCTTTG 833
Db        390 TTACCTTTTTATTTTATGTTATTTTATTTTATTTTATTTTTACTTAGTATATAATATTAT 449

Qy        834 ATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTTATATTAAAT 893
Db        450 ATTATATCATAGTATAGTTATATTATGGTGACTTTATTCATTATATAGATTGTATTTTGT 509

Qy        894 CATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTT 946
Db        510 GAACATAATATATATGCTATTTCTATTTCTATTTTATTTTATTTTATTTTATT 562

RESULT 6
US-09-252-991A-8972
; Sequence 8972, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 8972
;   LENGTH: 717
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8972

  Query Match             1.8%;  Score 63;  DB 4;  Length 717;
  Best Local Similarity   51.5%;  Pred. No. 7.4e-07;
  Matches  169;  Conservative    0;  Mismatches  156;  Indels    3;  Gaps    1;

Qy       3258 AGCGCGTTAGTAAGAACCGTGTCAGCATCACCTATGACGTTGAAACGAATGGCGCCGTAA 3317
Db        200 ACCGGGTGCGCCCGCCGCGCGTGCAGATCACCTACGACGTCGAGATCGGCAACGCCATCG 259

Qy       3318 AGACGAAAGAGCTGCCGTTTGTTGTTGGCGTCATTGGCGACTTTTCAGGACACAAACCAG 3377
Db        260 AGAAGAAGGAGCTGCCCCTGGTGGTCGGCATCCTCGCCGACCTCTCCGG---CAAGCCGG 316

Qy       3378 AATCAGAAAAAGTTGATTTAGAAGAGCGAGAGTTCACGGGTATCGATAAAGACAACTTCG 3437
Db        317 ATACCCCGCCGGCGAAGCTGGTCGAACGGCGTTTCGTCGACATCGACCGCGACAACTTCA 376

Qy       3438 ATACAGTGATGGGGCAAATTCACCCGCGTCTTTCGTACAAGGTTGATAACAAGCTTGCTA 3497
Db        377 ACGAGATCCTCTCCTCCATCTCTCCCCGCGCCACCCTGCAGGTCGACAACACCATCAGCG 436

Qy       3498 ATGATGATAGCCAGTTTGAAGTGAACTTGAGCCTCCGTTCGATGAAAGATTTCCACCCAG 3557
Db        437 GCGACGACAGCAAGCTGAACGTCGAGCTGCGCTTCAACCACATCGAGGACTTCGACCCGG 496

Qy       3558 AGAACTTAGTTGATNAAATTGAGCCGCT 3585
Db        497 TCAACCTGGTGAAGCAGGTGGTCCCGTT 524

RESULT 7
US-09-252-991A-8861
; Sequence 8861, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 8861
;   LENGTH: 819
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8861

  Query Match             1.8%;  Score 63;  DB 4;  Length 819;
  Best Local Similarity   51.5%;  Pred. No. 7.9e-07;
```

```
  Matches  169;  Conservative    0;  Mismatches  156;  Indels    3;  Gaps    1;

Qy       3258 AGCGCGTTAGTAAGAACCGTGTCAGCATCACCTATGACGTTGAAACGAATGGCGCCGTAA 3317
             | || ||  |   |  || ||    |||||||| ||||| || |     |||| |
Db         84 ACCGGGTGCGCCCGCCGCGCGTGCAGATCACCTACGACGTCGAGATCGGCAACGCCATCG 143

Qy       3318 AGACGAAAGAGCTGCCGTTTGTTGTTGGCGTCATTGGCGACTTTTCAGGACACAAACCAG 3377
             ||| ||| ||||||||  | || || ||| || | | |||| | || ||   ||| || |
Db        144 AGAAGAAGGAGCTGCCCCTGGTGGTCGGCATCCTCGCCGACCTCTCCGG---CAAGCCGG 200

Qy       3378 AATCAGAAAAAGTTGATTTAGAAGAGCGAGAGTTCACGGGTATCGATAAAGACAACTTCG 3437
             | |        |  | | | || ||    |||   | |||||    ||||||||| |
Db         201 ATACCCCGCCGGCGAAGCTGGTCGAACGGCGTTTCGTCGACATCGACCGCGACAACTTCA 260

Qy       3438 ATACAGTGATGGGGCAAATTCACCCGCGTCTTTCGTACAAGGTTGATAACAAGCTTGCTA 3497
             |    |  |       ||    || ||      |    |||| || ||||   |
Db         261 ACGAGATCCTCTCCTCCATCTCTCCCCGCGCCACCCTGCAGGTCGACAACACCATCAGCG 320

Qy       3498 ATGATGATAGCCAGTTTGAAGTGAACTTGAGCCTCCGTTCGATGAAAGATTTCCACCCAG 3557
                || || ||| || | | ||  |  || || ||     ||  | || ||| |||| |
Db         321 GCGACGACAGCAAGCTGAACGTCGAGCTGCGCTTCAACCACATCGAGGACTTCGACCCGG 380

Qy       3558 AGAACTTAGTTGATNAAATTGAGCCGCT 3585
                ||| | ||  |  |  | | |  ||| |
Db         381 TCAACCTGGTGAAGCAGGTGGTCCCGTT 408


RESULT 8
US-10-204-708-63
; Sequence 63, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 63
;   LENGTH: 5562
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-63

  Query Match             1.7%;  Score 60.4;  DB 4;  Length 5562;
  Best Local Similarity   48.0%;  Pred. No. 9.5e-06;
  Matches  172;  Conservative    0;  Mismatches  186;  Indels    0;  Gaps    0;

Qy        583 TGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTG 642
              | | | | || ||| | ||      | |    || || || |   | |||||| |   |
Db       1351 TTTGTTTGGGGTGATTTTTTTTTTGTTTTTTTTATTATTTTAATATTTTATTATTATTTT 1410

Qy        643 TATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCA 702
              | ||  |          ||| | | |    |  | |||||||  | ||     ||
Db       1411 TTTTTTTTCGGTTAGGAATGATTTTTTTTTTGATTATATTTATATTTTTTTATTTTAGTT 1470

Qy        703 AGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTAT 762
                ||  ||| | | |||| ||||| | |       | | |  |   ||  || |   || |

Db       1471 TAGTTTTTTTTTTTTTTATTTATTTTTTATTATTTTTATTTTTTGTTAGTTAGATTTTTT 1530

Qy        763 TTTTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGA 822
              |||| |  || |||  || ||  | |  |||   || |  |||  | |||| | | |  |
Db       1531 TTTTTTTATTATTTTTTTGTTTTTTATAGTGTTTTTTTTTATTTTTTTTTTGTTTAAGAA 1590

Qy        823 CAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTT 882
                  |    |  | | |||      || ||  |   | |||    ||     | ||
Db       1591 AGTTTATTAATTTTTATTTTTTTTATATTTTTTTTTTTTTTTAGGATAGTGTTATTTTTAG 1650

Qy        883 TTATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGT 940
                | |  || ||||| | |  | ||||  |||||||  | |||||||  | ||
Db       1651 AATTTTGTGATAATAATGATTTTTATTATATTATTTATTTTTGTTTTTTTTATTTTAT 1708


RESULT 9
US-10-204-708-20
; Sequence 20, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 20
;   LENGTH: 6866
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-20

  Query Match             1.7%;  Score 60;  DB 4;  Length 6866;
  Best Local Similarity   44.7%;  Pred. No. 1.3e-05;
  Matches  287;  Conservative    0;  Mismatches  345;  Indels   10;  Gaps    1;

Qy        495 CTTAATGGTGAGGGAGCGGGTTTGGTGCTTTAATTCGTTACTGCTCATATTCAATTAATT 554
              | | ||| ||      |  ||||| |  |||  | | | |  |    || ||| ||
Db       5366 CGTTATGTTGGTTAGGTTGGTTTTGAATTTTTGATTTTAAGTAATTTGTTTGTTTTAGTT 5425

Qy        555 CACTATTAAATAAACAGTTCTAAAAGGCTGTTTATTGGATGAATATTCGAAATTATCACA 614
                ||      | | | |      | |  | |  |   |   |   |  | |  |
Db       5426 TTTTAAAGTGTTGAGATTATAGGTATGAGTTATTATTTTTAGTTTAAAGTTTGTTTTTTA 5485

Qy        615 TAATAATTGATGCTATTATTACTTGCTGTATTGGTATCAACTTTCATGCTCTATACATGT 674
              || |  ||   | | ||  |  ||  | | ||   |   | ||| |      ||    |
Db       5486 TATTTTTTATGGTTTTTGATGTTTTTTTTGTTTTAAATTATTTTTAATGAGGATTTTAAT 5545

Qy        675 AATATATTTCGAGTTAGACCTTAATTCAAGGTAATTTGTCTATTTAATTATTATCTGAAT 734
              || |  || ||  ||  |  ||  ||      |  ||| |  |||| || ||  |   |
Db       5546 TATTTTATTAGAATTTAAATTTTTTTATTATTTTTTTTTTTAATTTGGTTTTTTGTTTTGT 5605

Qy        735 AATATGTAATCGATTGCTTTGTGGTTATTTTTATGTTTGTTTCATTTTTAATGACGGTGA 794
              | | |   | |||  | |   || | ||| ||||  |  |  |||  |||  |
Db       5606 ATTTTTGTTTTTATTTTTATAGTTTTTTATTTTTGTTTTATAGAAGTTTTTTGATTTTAT 5665

Qy        795 GCTTGTGCATTCATATTTTTTATGATGACAACATCTTTGATGAAGTATTTAAGATATTGT 854
              | || |   || || | ||||  ||      |  | |    |  | ||| | |||||
```

```
Db       5666 GTTTATTTTTTTATTTATTTTTATATAGATTTAGATATATGTAGATTTTTTATATATTTA 5725

Qy        855 TAATGCATGAGGGGTTTGCGTGTATTTTTTATATTAAATCATAATAAAATCAACAATATA 914

Db       5726 TTTAGATATATTTAGATATAAATATATATTATTTTTATTTAAATTATAGATATAGATAGA 5785

Qy        915 TGTTATTTTGTGTCTTTTTATAGTGTTCTTTTAAAGAGGTAGGATGACCTAAAGGTCGCC 974

Db       5786 TAGAAATTTATAGTTGGTTTTTTGTGGGTTTTGTATTTGTGGATTTAATTAATTGTGGTT 5845

Qy        975 TAAATATGGCGTAAATTGCCATTGCTATAATTCACCTCAAAGATACACTATTGGCAAATT 1034

Db       5846 TGAAAATATCGAAAA----------GAAAAATTAATTAAAAAATAATAATATAGTAATAA 5895

Qy       1035 GACAAATATGTCACTTCGTATGAAACAATATTAGTAGATGTTGTTTTTGCTGCAAAAATA 1094

Db       5896 AATAATTTATAAAATTAGAAAAGTTAAATTTTATTTGAGGTTTTTTTAATATTTTATTTA 5955

Qy       1095 AAAATTTTTCTGGTTGAAATAACTCAAGGCCTCTAGCGTTTT 1136

Db       5956 AAATTTGATATTTTTGTTTTAATTTATAATATATAGGGATTT 5997


RESULT 10
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             1.7%;  Score 59.8;  DB 4;  Length 640681;
  Best Local Similarity   44.1%;  Pred. No. 0.00013;
  Matches  250;  Conservative    0;  Mismatches  317;  Indels    0;  Gaps    0;

Qy        515 TTTGGTGCTTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTC 574

Db     325026 TTTGATATTTAAAATTAATAATTTTTTTGTTTTTTAAAATTTTATTTTTTGTAATTTTCA 325085

Qy        575 TAAAAGGCTGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATT 634

Db     325086 TAGAACGTATTTCTTTAAATAAAGTAATAATTTTTTAATTTAAATTAAACTAATTTTTAT 325145

Qy        635 ACTTGCTGTATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACC 694

Db     325146 TCATATTATATTTTTTATAAATTTAAAAATATTTCTTAAATAAATATTTTAAATGTGTAT 325205

Qy        695 TTAATTCAAGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTT 754

Db     325206 TTAAAAAAATATAAATAATAATTATTTTATTTTACATTTTATTATATTAAAATTTATACT 325265

Qy        755 GTGGTTATTTTTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTT 814

Db     325266 ATAAATATTTTTATATTTTAATAAAAATACATTTTATATAAATAAAATACTGATATTTTT 325325

Qy        815 TATGATGACAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCG 874

Db     325326 TCTTTTTATTTTTTTAAATTACTATAATGATAATTATAGTATATAATAAAATAAGTAAAA 325385

Qy        875 TGTATTTTTTATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTA 934

Db     325386 TAAAGAACAAAAAAATTATCAATATAAATAAATAAGGATGTAATTTCTTTTTTTTTATTC 325445

Qy        935 TAGTGTTCTTTTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCC 994

Db     325446 TAATTTTATATAAAAAAAGAATTTGCATCGAATCGATAAAATAAAAATTTTTATTTTCAA 325505

Qy        995 ATTGCTATAATTCACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTA 1054

Db     325506 TTGAATATATTCAAAACATTCAATAAAAACTTTATAAATTTAAAATGACATCAGAATGTA 325565

Qy       1055 TGAAACAATATTAGTAGATGTTGTTTT 1081

Db     325566 TAAACCTATCTCTCTTTTTATTGCTTT 325592


RESULT 11
US-09-252-991A-752
; Sequence 752, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 752
;   LENGTH: 600
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-752

  Query Match             1.7%;  Score 59.4;  DB 4;  Length 600;
  Best Local Similarity   56.1%;  Pred. No. 6.1e-06;
  Matches  133;  Conservative    0;  Mismatches  101;  Indels    3;  Gaps    1;

Qy       3275 CGTGTCAGCATCACCTATGACGTTGAAACGAATGGCGCCGTAAAGACGAAAGAGCTGCCG 3334

Db        133 CGCGTGCAGATCGAGTACGACGTGGAGTTGTACGGTGCCGAGAAGAAGGTCCAGCTGCCC 192

Qy       3335 TTTGTTGTTGGCGTCATTGGCGACTTTTCAGGACACAAACCAGAATCAGAAAAAGTTGAT 3394

Db        193 TTCGTCATGGGCGTCATGGCCGACCTCGCCGG---CAAGCCCGCCGAACCCCAGGCGGCG 249

Qy       3395 TTAGAAGAGCGAGAGTTCACGGGTATCGATAAAGACAACTTCGATACAGTGATGGGGCAA 3454

Db        250 GTCGCCGACCGCAAGTTCCTGGAGATCGACGTGGACAACTTCGATGCCCGGCTGAAGGCG 309

Qy       3455 ATTCACCCGCGTCTTTCGTACAAGGTTGATAACAAGCTTGCTAATGATGATAGCCAG 3511

Db        310 ATGAAGCCGCGGGTGGCCTTCAACGTACCGAACGTGCTGACCGGCGAAGGCAACCTG 366


RESULT 12
US-09-252-991A-723
; Sequence 723, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
```

; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 723
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-723

Query Match 1.7%; Score 59.4; DB 4; Length 681;
Best Local Similarity 56.1%; Pred. No. 6.5e-06;
Matches 133; Conservative 0; Mismatches 101; Indels 3; Gaps 1;

Qy 3275 CGTGTCAGCATCACCTATGACGTTGAAACGAATGGCGCCGTAAAGACGAAAGAGCTGCCG 3334
Db 140 CGCGTGCAGATCGAGTACGACGTGGAGTTGTACGGTGCCGAGAAGAAGGTCCAGCTGCCC 199

Qy 3335 TTTGTTGTTGGCGTCATTGGCGACTTTTCAGGACACAAACCAGAATCAGAAAAAGTTGAT 3394
Db 200 TTCGTCATGGGCGTCATGGCCGACCTCGCCGG---CAAGCCCGCCGAACCCCAGGCGGCG 256

Qy 3395 TTAGAAGAGCGAGAGTTCACGGGTATCGATAAAGACAACTTCGATACAGTGATGGGGCAA 3454
Db 257 GTCGCCGACCGCAAGTTCCTGGAGATCGACGTGGACAACTTCGATGCCCGGCTGAAGGCG 316

Qy 3455 ATTCACCCGCGTCTTTCGTACAAGGTTGATAACAAGCTTGCTAATGATGATAGCCAG 3511
Db 317 ATGAAGCCGCGGGTGGCCTTCAACGTACCGAACGTGCTGACCGGCGAAGGCAACCTG 373

RESULT 13
US-09-252-991A-658/c
; Sequence 658, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 658
; LENGTH: 2019
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-658

Query Match 1.7%; Score 59.4; DB 4; Length 2019;
Best Local Similarity 56.1%; Pred. No. 1.1e-05;
Matches 133; Conservative 0; Mismatches 101; Indels 3; Gaps 1;

Qy 3275 CGTGTCAGCATCACCTATGACGTTGAAACGAATGGCGCCGTAAAGACGAAAGAGCTGCCG 3334
Db 1935 CGCGTGCAGATCGAGTACGACGTGGAGTTGTACGGTGCCGAGAAGAAGGTCCAGCTGCCC 1876

Qy 3335 TTTGTTGTTGGCGTCATTGGCGACTTTTCAGGACACAAACCAGAATCAGAAAAAGTTGAT 3394
Db 1875 TTCGTCATGGGCGTCATGGCCGACCTCGCCGG---CAAGCCCGCCGAACCCCAGGCGGCG 1819

Qy 3395 TTAGAAGAGCGAGAGTTCACGGGTATCGATAAAGACAACTTCGATACAGTGATGGGGCAA 3454
Db 1818 GTCGCCGACCGCAAGTTCCTGGAGATCGACGTGGACAACTTCGATGCCCGGCTGAAGGCG 1759

Qy 3455 ATTCACCCGCGTCTTTCGTACAAGGTTGATAACAAGCTTGCTAATGATGATAGCCAG 3511
Db 1758 ATGAAGCCGCGGGTGGCCTTCAACGTACCGAACGTGCTGACCGGCGAAGGCAACCTG 1702

RESULT 14
US-08-883-795A-36
; Sequence 36, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcuve, Genevieve
; APPLICANT: Awang, Gregor
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,795A
; FILING DATE: 27-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 7841-062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 665 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: Rh 32
US-08-883-795A-36

Query Match 1.6%; Score 59.2; DB 2; Length 665;
Best Local Similarity 45.8%; Pred. No. 7.2e-06;
Matches 242; Conservative 0; Mismatches 283; Indels 3; Gaps 1;

Qy 515 TTTGGTGCTTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTC 574
Db 112 TTTTATAATTAAAATATTTATAATTAAATATTTTATAATTAAAATATTTATAATTAAATA 171

Qy 575 TAAAAGGCTGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATT 634
Db 172 TTTTATAATTAAAATATTTATAATTAAATATTTTATAATTAAAATATTTATAATTAAATA 231

Qy 635 ACTTGCTGTATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACC 694
Db 232 TTTTATAATTAAAATATTTATAATTAAATATTTTATAATTAAAATGTTTATAATTAAATA 291

Qy 695 TTAATTCAAGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTT 754
Db 292 TTTTATAATTAAAATGTTTATAATTACATATTTTATAATTAAAATGTTTATAATTACATA 351

Qy 755 GTGGTTATTTTTA---TGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATT 811
Db 352 TTTTATAATTAAAATGTTTATAATTACATATTTTATAATTAAAATGTTTATAATTACATA 411

Qy 812 TTTTATGATGACAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTT 871

```
Db        412 TTTTATAATTAAAATGTTTATAATTACATATTTTATAATTACATATTTTATAAAGTATTT 471

Qy        872 GCGTGTATTTTTTATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTT 931
              ||  | || ||| |    |  ||  |       |||| |  || ||   || |||
Db        472 ATAATTACATATTTTATAATTAAAGTATTTATAATTACATATTTTATAATTAAAGTATTT 531

Qy        932 TTATAGTGTTCTTTTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATT 991
              ||       | ||||| |      |   |      ||  ||      | |   ||   |||||
Db        532 ATAATTACATATTTTATAATTCAATATTTTATAAATAGTTAAAAAGACGAGGAAAAAATT 591

Qy        992 GCCATTGCTATAATTCACCTCAAAGATACACTATTGGCAAATTGACAA 1039
                |  | |   |     ||  ||  |   |||| || || ||| ||
Db        592 AAAAAGACGAGGTTATTGATCTCAGGAATTGTATTTGCCAAGTGAGAA 639
```

RESULT 15
US-10-204-708-55
; Sequence 55, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 55
; LENGTH: 11015
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-55

Query Match 1.6%; Score 57.6; DB 4; Length 11015;
Best Local Similarity 49.7%; Pred. No. 7.2e-05;
Matches 173; Conservative 0; Mismatches 174; Indels 1; Gaps 1;

```
Qy        585 TTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTA 644
              ||| | |  ||  ||     |  ||||     | ||   | ||| |    ||  ||    |
Db      10659 TTTTTAGAGTGGTTAAAGTATTTTATATTTTTATTGATAATGTTTAATTTTTTTATATTT 10718

Qy        645 TTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAG 704
              ||  ||  |  | | ||  | | |   | |  |||  ||  | |||    ||| || ||
Db      10719 TTAATAATATTTGTTATTTTATTTGTTTTTTTTATTATTTTATTTAATTATTATTTTAAT 10778

Qy        705 GTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTT 764
              |    | |  |  |   |||||||     | || |||  |   ||  |   |  | || |
Db      10779 GGGTATGGAGTGGTATTTTATTATGGTTTTGATTTGTGTTTTTTTAATGATTAATAATAT 10838

Qy        765 TTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACA 824
              | |   || ||| || | |  ||  ||| |  || |   ||  |    | ||| ||
Db      10839 TGAGTATTTTTTTATATGT-TTGTTGGTTATTTTTTATTTTTTTTGGAGTAATGTTGGTT 10897

Qy        825 ACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTT 884
              | || |||| |    || |||||| ||| ||  |  ||   || || |   | ||| | |
Db      10898 ATATTTTTGGTATTATAATTAAGAAATTATTGTTAAATTTAGGATTAGAAAGAATTATAT 10957

Qy        885 ATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTT 932
              | |  || | ||  |   |     ||| ||| || |||  |   ||||
```

```
Db      10958 AAAGGAATTTATTGTTTTAAACGTAATTTATATTTTTTATTTATTTTT 11005
```

Search completed: July 2, 2004, 22:59:43
Job time : 244 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 2, 2004, 20:30:55 ; Search time 1494 Seconds
(without alignments)
11585.848 Million cell updates/sec

Title: US-10-780-347-1
Perfect score: 3588
Sequence: 1 gtcgacttattgcattgatg.....gatnaaattgagccgcttaa 3588

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3163042 seqs, 2412103800 residues

Total number of hits satisfying chosen parameters: 6326084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	72.8	2.0	11422	15	US-10-311-455-192	Sequence 192, App
2	72.8	2.0	11422	17	US-10-257-166-18	Sequence 18, Appl
3	72.8	2.0	3673778	15	US-10-312-841-2	Sequence 2, Appli
4	71.4	2.0	61020	13	US-10-221-714A-514	Sequence 514, App
5	69.4	1.9	8093	13	US-10-221-714A-158	Sequence 158, App
6	69.4	1.9	8093	15	US-10-172-086-26	Sequence 26, Appl
7	69.4	1.9	8093	15	US-10-204-708-32	Sequence 32, Appl
8	69.4	1.9	8093	17	US-10-240-589C-46	Sequence 46, Appl
9	69.4	1.9	8093	17	US-10-311-507-58	Sequence 58, Appl
10	69	1.9	5504	13	US-10-221-613-270	Sequence 270, App
11	68	1.9	6306	15	US-10-239-676-224	Sequence 224, App
12	68	1.9	6306	15	US-10-240-453-332	Sequence 332, App
13	68	1.9	13131	17	US-10-240-589C-58	Sequence 58, Appl
14	66.6	1.9	8666	13	US-10-221-714A-27	Sequence 27, Appl
15	66.6	1.9	8666	15	US-10-311-455-369	Sequence 369, App
16	66.6	1.9	8666	17	US-10-311-507-95	Sequence 95, Appl
17	66.6	1.9	9504	15	US-10-240-453-281	Sequence 281, App
18	66.2	1.8	17848	15	US-10-239-676-28	Sequence 28, Appl
19	66.2	1.8	17848	15	US-10-240-453-38	Sequence 38, Appl
20	66.2	1.8	17848	17	US-10-257-166-58	Sequence 58, Appl
21	66.2	1.8	18624	15	US-10-311-455-1675	Sequence 1675, Ap
22	66	1.8	516	9	US-09-960-352-5785	Sequence 5785, Ap
23	65.8	1.8	7849	17	US-10-240-589C-88	Sequence 88, Appl
24	65.8	1.8	3673778	15	US-10-312-841-1	Sequence 1, Appli
25	65.2	1.8	6040	15	US-10-204-708-69	Sequence 69, Appl
26	65.2	1.8	6040	17	US-10-240-589C-109	Sequence 109, App
27	65	1.8	9539	15	US-10-239-676-52	Sequence 52, Appl
28	65	1.8	9539	15	US-10-240-453-54	Sequence 54, Appl
29	65	1.8	9642	15	US-10-311-455-330	Sequence 330, App
30	65	1.8	12237	15	US-10-311-455-2331	Sequence 2331, Ap
31	64.8	1.8	9905	15	US-10-311-455-35	Sequence 35, Appl
32	64.8	1.8	11555	15	US-10-311-455-589	Sequence 589, App
33	64.8	1.8	18183	13	US-10-221-714A-1	Sequence 1, Appli
34	64.6	1.8	6233	13	US-10-221-714A-219	Sequence 219, App
35	64.6	1.8	7571	15	US-10-311-455-500	Sequence 500, App
36	64.6	1.8	8079	17	US-10-240-589C-122	Sequence 122, App
37	64.4	1.8	11670	15	US-10-240-452-25	Sequence 25, Appl
38	64.2	1.8	7667	13	US-10-221-714A-56	Sequence 56, Appl
39	64.2	1.8	15743	15	US-10-240-453-269	Sequence 269, App
40	64	1.8	12705	15	US-10-311-455-121	Sequence 121, App
41	63.8	1.8	6419	15	US-10-311-455-240	Sequence 240, App
42	63.6	1.8	7810	15	US-10-239-676-142	Sequence 142, App
43	63.6	1.8	7810	15	US-10-240-453-156	Sequence 156, App
44	63.4	1.8	17893	15	US-10-311-455-1338	Sequence 1338, Ap
45	63.2	1.8	5750	13	US-10-221-714A-433	Sequence 433, App

ALIGNMENTS

RESULT 1
US-10-311-455-192
; Sequence 192, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 192
; LENGTH: 11422
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-192

Query Match 2.0%; Score 72.8; DB 15; Length 11422;
Best Local Similarity 47.7%; Pred. No. 5.8e-06;
Matches 244; Conservative 0; Mismatches 267; Indels 1; Gaps 1;

```
Qy       605 AATTATCACATAATAATTGATGCTATTATTACTTGCTGTATTGGTATCAACTTTCATGCT 664
             |||||     | | |||   ||    |  | | | | ||| | ||   |   |  |
Db      9660 AATTAATTTTTTAAAATATGTGTATGTGGTTATAGTTATATAGATAAATATAGTGGTAGT 9719

Qy       665 CTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAGGTAATTTGTCTATTTAATTA 724
```

```
Db       9720 TAAATGGTTTATTATTATTTTTTTTGTTTTTTATTTTATGATATTAATTTTTTATTAAAA 9779

Qy        725 TTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTTTATGTTTGTTTCATTTTTA 784

Db       9780 TATTTTGATTATTGTGTGATATATTATTATATGGGTATATTATGGTTTATTGTATTTTTT 9839

Qy        785 ATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAACATCTTTGATGAAGTATTT 844

Db       9840 GTTATTTATATGATGATAATAATAATTTTAGGTAATAAATATTATTTGAAAGTATTATTT 9899

Qy        845 AAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTT-TTTATATTAAATCATAATAAAA 903

Db       9900 TATTTATTTTGAATATATGAATTTATTTAAAATTTATGATTATTTTTAAAATAAGTATTA 9959

Qy        904 TCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTTAAAGAGGTAGGATGACC 963

Db       9960 TTATTAATTTTATTTTATATTTGAGAAAATATAGTTATAAAGAGTTTAAGTAATTTGTTT 10019

Qy        964 TAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAATTCACCTCAAAGATACACT 1023

Db      10020 TAAATTATATAGATATCGAGGTTAGTTTTTGATTTTTTTTATGTATTATTATAATAGGTA 10079

Qy       1024 ATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATATTAGTAGATGTTGTTTTTG 1083

Db      10080 TTTTGTATATCGATATTTGAGTGTATTTTTTAATTATTTTTTTAATTTATTTTTTTAGAA 10139

Qy       1084 CTGCAAAAATAAAAATTTTTCTGGTTGAAATA 1115

Db      10140 TTGGAATAATTAAATACGATATGGATAAAGTA 10171
```

RESULT 2
US-10-257-166-18
; Sequence 18, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 18
; LENGTH: 11422
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-18

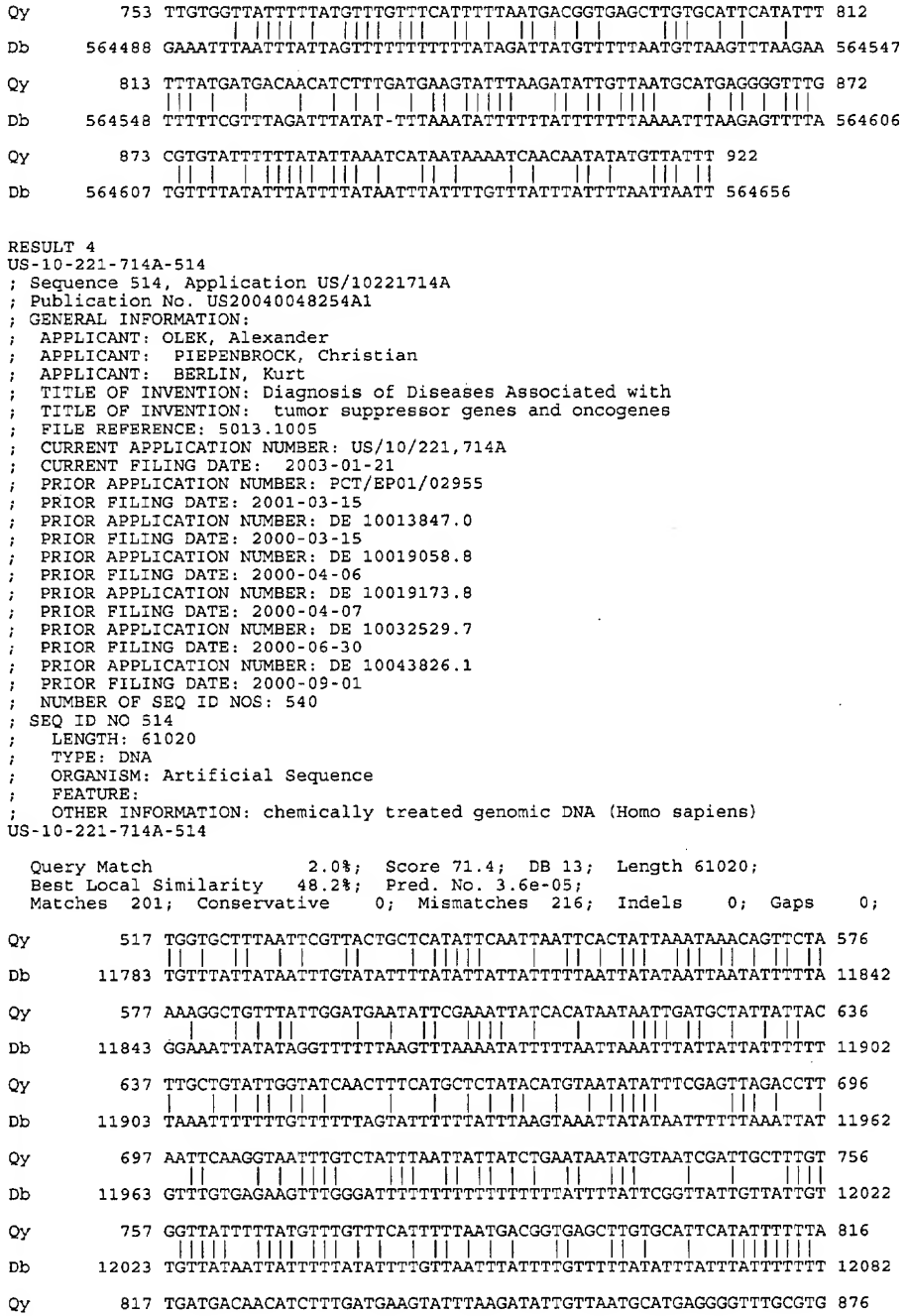
Query Match 2.0%; Score 72.8; DB 17; Length 11422;
Best Local Similarity 47.7%; Pred. No. 5.8e-06;
Matches 244; Conservative 0; Mismatches 267; Indels 1; Gaps 1;

```
Qy        605 AATTATCACATAATAATTGATGCTATTATTACTTGCTGTATTGGTATCAACTTTCATGCT 664

Db       9660 AATTAATTTTTTAAAATATGTGTATGTGGTTATAGTTATATAGATAAATATAGTGGTAGT 9719

Qy        665 CTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAGGTAATTTGTCTATTTAATTA 724

Db       9720 TAAATGGTTTATTATTATTTTTTTTGTTTTTTATTTTATGATATTAATTTTTTATTAAAA 9779

Qy        725 TTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTTTATGTTTGTTTCATTTTTA 784

Db       9780 TATTTTGATTATTGTGTGATATATTATTATATGGGTATATTATGGTTTATTGTATTTTTT 9839

Qy        785 ATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAACATCTTTGATGAAGTATTT 844

Db       9840 GTTATTTATATGATGATAATAATAATTTTAGGTAATAAATATTATTTGAAAGTATTATTT 9899

Qy        845 AAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTT-TTTATATTAAATCATAATAAAA 903

Db       9900 TATTTATTTTGAATATATGAATTTATTTAAAATTTATGATTATTTTTAAAATAAGTATTA 9959

Qy        904 TCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTTAAAGAGGTAGGATGACC 963

Db       9960 TTATTAATTTTATTTTATATTTGAGAAAATATAGTTATAAAGAGTTTAAGTAATTTGTTT 10019

Qy        964 TAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAATTCACCTCAAAGATACACT 1023

Db      10020 TAAATTATATAGATATCGAGGTTAGTTTTTGATTTTTTTTATGTATTATTATAATAGGTA 10079

Qy       1024 ATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATATTAGTAGATGTTGTTTTTG 1083

Db      10080 TTTTGTATATCGATATTTGAGTGTATTTTTTAATTATTTTTTTAATTTATTTTTTTAGAA 10139

Qy       1084 CTGCAAAAATAAAAATTTTTCTGGTTGAAATA 1115

Db      10140 TTGGAATAATTAAATACGATATGGATAAAGTA 10171
```

RESULT 3
US-10-312-841-2
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (379615)
US-10-312-841-2

Query Match 2.0%; Score 72.8; DB 15; Length 3673778;
Best Local Similarity 50.0%; Pred. No. 0.00022;
Matches 235; Conservative 0; Mismatches 232; Indels 3; Gaps 2;

```
Qy        455 CGATTAATTGCGTGACAATATAACTTTTGCTATCTGAAAGCTTAATGGTGAGGGAGCGGG 514

Db     564188 CGATTATTTTATTGTGGTTTTAATTTGTATTTTTTTAATGGTTAATGATGTTGAATAATT 564247

Qy        515 TTTGGTGCTTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTC 574

Db     564248 TTTTATGTGTTAATTTGTTATTTTTATATTTTTTTTAGTGAAATATTATTTTAAGTTTTT 564307

Qy        575 TAAAAGGCTGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATT 634

Db     564308 TGTTTGTTTTTTAATTGGATTTTTAATGTGTTTTTTTTGTTGAGTTTTGAGAGTTTTTTA 564367

Qy        635 ACTTGCTGTATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACC 694

Db     564368 TATATTTTAGATATTAGTTTTTTGTTTAATGTATAGTTTGTAAATATTTTTTTTTTATTGT 564427

Qy        695 TTAATTCA--AGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCT 752

Db     564428 GTAGTTTATTTTTTTTATTTATTTATTAGATTTTTATAAAGTAATAGTTTTATATTTTGAT 564487
```

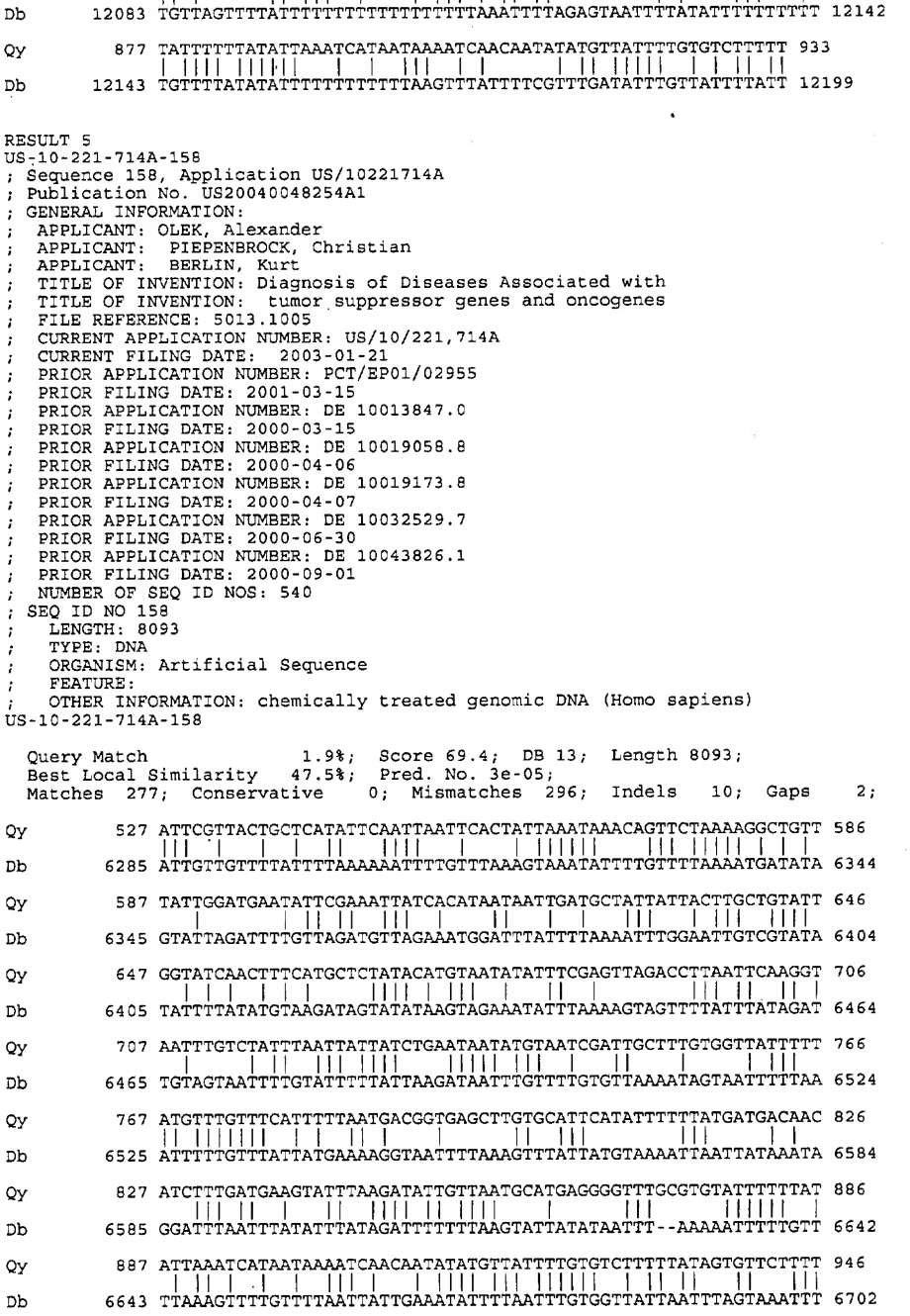


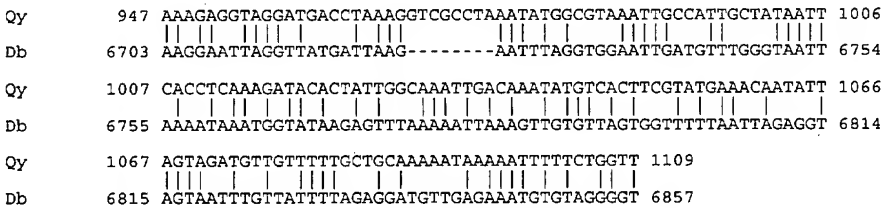
RESULT 4
US-10-221-714A-514
; Sequence 514, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 514
; LENGTH: 61020
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-514

Query Match 2.0%; Score 71.4; DB 13; Length 61020;
Best Local Similarity 48.2%; Pred. No. 3.6e-05;
Matches 201; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

RESULT 5
US-10-221-714A-158
; Sequence 158, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 158
; LENGTH: 8093
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-158

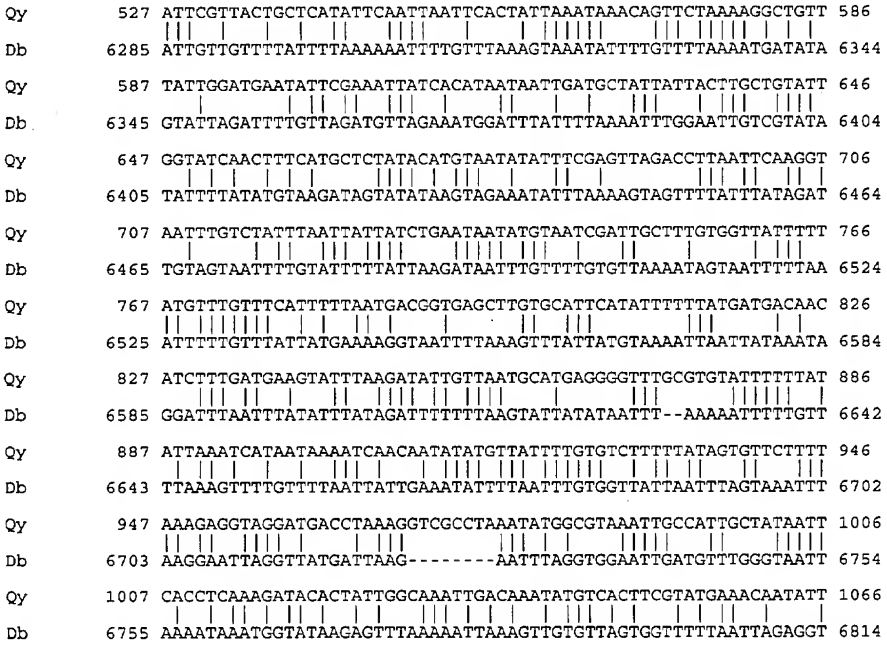
Query Match 1.9%; Score 69.4; DB 13; Length 8093;
Best Local Similarity 47.5%; Pred. No. 3e-05;
Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;

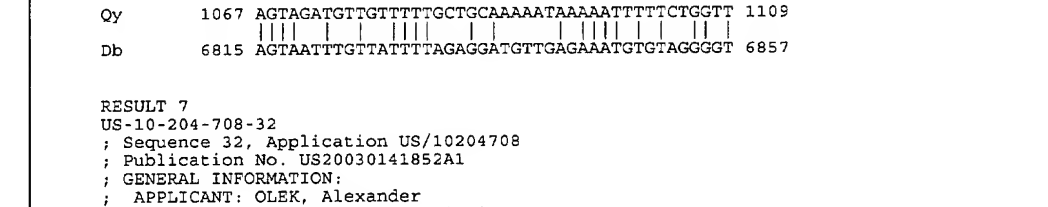




RESULT 6
US-10-172-086-26
; Sequence 26, Application US/10172086
; Publication No. US20030113750A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Method and nucleic acids for the differentiation
; TITLE OF INVENTION: of prostate tumors
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/172,086
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 26
; LENGTH: 8093
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-172-086-26

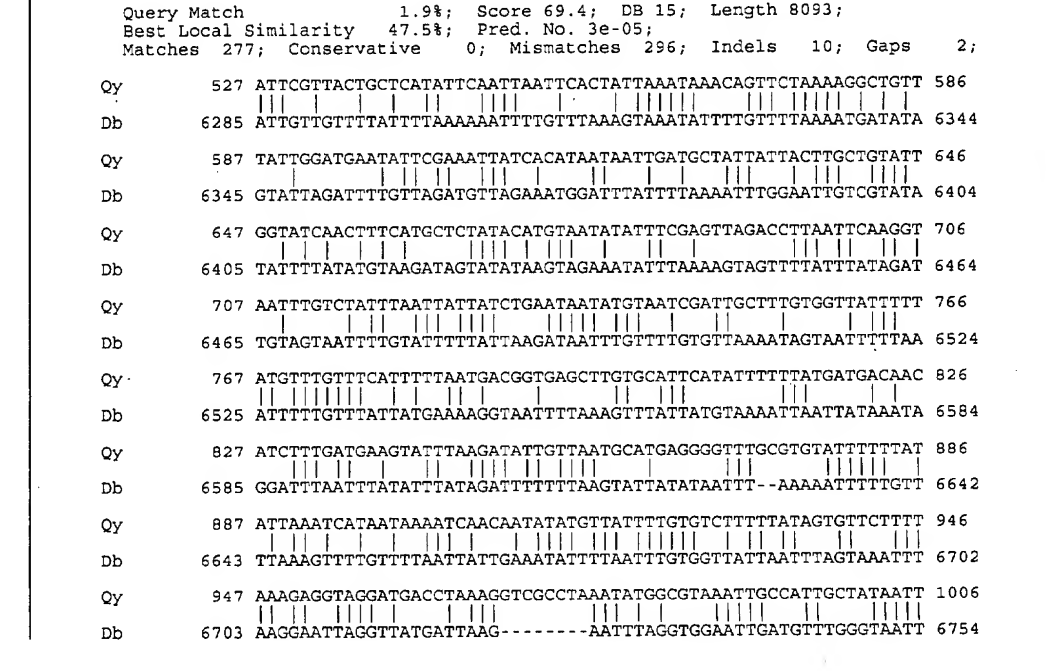
Query Match 1.9%; Score 69.4; DB 15; Length 8093;
Best Local Similarity 47.5%; Pred. No. 3e-05;
Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;





RESULT 7
US-10-204-708-32
; Sequence 32, Application US/10204708
; Publication No. US20030141852A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 32
; LENGTH: 8093
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-32

Query Match 1.9%; Score 69.4; DB 15; Length 8093;
Best Local Similarity 47.5%; Pred. No. 3e-05;
Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;



```
Qy       1007 CACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATATT 1066
Db       6755 AAAATAAATGGTATAAGAGTTTAAAAATTAAAGTTGTGTTAGTGGTTTTTAATTAGAGGT 6814

Qy       1067 AGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTT 1109
Db       6815 AGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTGTAGGGGT 6857
```

RESULT 8
US-10-240-589C-46
; Sequence 46, Application US/10240589C
; Publication No. US20040076956A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: DNA repair
; FILE REFERENCE: 5013.1008
; CURRENT APPLICATION NUMBER: US/10/240,589C
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03972
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 148
; SEQ ID NO 46
; LENGTH: 8093
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-589C-46

Query Match 1.9%; Score 69.4; DB 17; Length 8093;
Best Local Similarity 47.5%; Pred. No. 3e-05;
Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;

```
Qy        527 ATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGCTGTT 586
Db       6285 ATTGTTGTTTTATTTTAAAAAATTTTGTTTAAAGTAAATATTTTGTTTTAAAATGATATA 6344

Qy        587 TATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTATT 646
Db       6345 GTATTAGATTTTGTTAGATGTTAGAAATGGATTTATTTTAAAATTTGGAATTGTCGTATA 6404

Qy        647 GGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAGGT 706
Db       6405 TATTTTATATGTAAGATAGTATATAAGTAGAAATATTTAAAAGTAGTTTTATTTATAGAT 6464

Qy        707 AATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTTT 766
Db       6465 TGTAGTAATTTTGTATTTTTATTAAGATAATTTGTTTTGTGTTAAAATAGTAATTTTTAA 6524

Qy        767 ATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAAC 826
Db       6525 ATTTTTGTTTATTATGAAAAGGTAATTTTAAAGTTTATTATGTAAAATTAATTATAAATA 6584

Qy        827 ATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTTAT 886
Db       6585 GGATTTAATTTATATTTATAGATTTTTTTAAGTATTATATAATTT--AAAAATTTTTGTT 6642

Qy        887 ATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTT 946
Db       6643 TTAAAGTTTTGTTTTAATTATTGAAATATTTTAATTTGTGGTTATTAATTTAGTAAATTT 6702
```

```
Qy        947 AAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAATT 1006
Db       6703 AAGGAATTAGGTTATGATTAAG--------AATTTAGGTGGAATTGATGTTTGGGTAATT 6754

Qy       1007 CACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATATT 1066
Db       6755 AAAATAAATGGTATAAGAGTTTAAAAATTAAAGTTGTGTTAGTGGTTTTTAATTAGAGGT 6814

Qy       1067 AGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTT 1109
Db       6815 AGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTGTAGGGGT 6857
```

RESULT 9
US-10-311-507-58
; Sequence 58, Application US/10311507
; Publication No. US20040115630A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and nucleic acids for the analysis of astrocytomas
; FILE REFERENCE: 5013.1013
; CURRENT APPLICATION NUMBER: US/10/311,507
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07538
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 58
; LENGTH: 8093
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-507-58

Query Match 1.9%; Score 69.4; DB 17; Length 8093;
Best Local Similarity 47.5%; Pred. No. 3e-05;
Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;

```
Qy        527 ATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGCTGTT 586
Db       6285 ATTGTTGTTTTATTTTAAAAAATTTTGTTTAAAGTAAATATTTTGTTTTAAAATGATATA 6344

Qy        587 TATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTATT 646
Db       6345 GTATTAGATTTTGTTAGATGTTAGAAATGGATTTATTTTAAAATTTGGAATTGTCGTATA 6404

Qy        647 GGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAGGT 706
Db       6405 TATTTTATATGTAAGATAGTATATAAGTAGAAATATTTAAAAGTAGTTTTATTTATAGAT 6464

Qy        707 AATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTTT 766
Db       6465 TGTAGTAATTTTGTATTTTTATTAAGATAATTTGTTTTGTGTTAAAATAGTAATTTTTAA 6524

Qy        767 ATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAAC 826
Db       6525 ATTTTTGTTTATTATGAAAAGGTAATTTTAAAGTTTATTATGTAAAATTAATTATAAATA 6584

Qy        827 ATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTTAT 886
Db       6585 GGATTTAATTTATATTTATAGATTTTTTTAAGTATTATATAATTT--AAAAATTTTTGTT 6642

Qy        887 ATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTT 946
Db       6643 TTAAAGTTTTGTTTTAATTATTGAAATATTTTAATTTGTGGTTATTAATTTAGTAAATTT 6702
```

```
Qy        947 AAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAATT 1006
Db       6703 AAGGAATTAGGTTATGATTAAG--------AATTTAGGTGGAATTGATGTTTGGGTAATT 6754

Qy       1007 CACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATATT 1066
Db       6755 AAAATAAATGGTATAAGAGTTTAAAAATTAAAGTTGTGTTAGTGGTTTTTAATTAGAGGT 6814

Qy       1067 AGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTT 1109
Db       6815 AGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTGTAGGGGT 6857
```

RESULT 10
US-10-221-613-270
; Sequence 270, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, ALexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221,613
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; DE 10013847.00
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-03-15
; 2000-03-15
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 270
; LENGTH: 5504
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-270

```
  Query Match             1.9%;  Score 69;  DB 13;  Length 5504;
  Best Local Similarity   46.2%;  Pred. No. 3e-05;
  Matches  265;  Conservative    0;  Mismatches  305;  Indels    3;  Gaps    1;

Qy        615 TAATAATTGATGCTATTATTACTTGCTGTATTGGTATCAACTTTCATGCTCTATACATGT 674
Db       1009 TTATAATTAGTGTTTGTTTTTTGTTATTTTTAGGTTTTTATTTTCGTTTTTTTTTTTTTT 1068

Qy        675 AATATATTTCGAGTTAGACCTTAATTCAAGGTAATTTGTCTATTTAATTATTATCTGAAT 734
Db       1069 TTTATATTTAGCGTT---TTTTTTTTTTTATTTTTTTTTTTTATTTTTTTTATTTGTTTTT 1125

Qy        735 AATATGTAATCGATTGCTTTGTGGTTATTTTTATGTTTGTTTCATTTTTAATGACGGTGA 794
Db       1126 TTTAGGATTATTTTTTTTTAGGATTAAATTTGATCGTAGTATATTTTTTTAGGTCGTTTG 1185

Qy        795 GCTTGTGCATTCATATTTTTTATGATGACAACATCTTTGATGAAGTATTTAAGATATTGT 854
Db       1186 TATTTTTTTTTGAAGTTTTTTATTAGAGAGTTTTTGTTTTTTTTATTTTAAAGTATTTAT 1245

Qy        855 TAATGCATGAGGGGTTTGCGTGTATTTTTTATATTAAATCATAATAAAATCAACAATATA 914
Db       1246 TTGTTTCGGTGGTATTTTTTGATATTAGTAAGAGAAAAAGATTTTAAAAGGATTGTTAGT 1305

Qy        915 TGTTATTTTGTGTCTTTTTATAGTGTTCTTTTAAAGAGGTAGGATGACCTAAAGGTCGCC 974
Db       1306 TTGATTATTTTTTTTTTTAATTATTTTTTTTTTTTTTTTGTATAGGTTTGTATTTTTTT 1365

Qy        975 TAAATATGGCGTAAATTGCCATTGCTATAATTCACCTCAAAGATACACTATTGGCAAATT 1034
Db       1366 TAAATTTTATGTAAATAGCGTTAACGATAAATATAATTATTATTAAGTGTTTAATTTATA 1425

Qy       1035 GACAAATATGTCACTTCGTATGAAACAATATTAGTAGATGTTGTTTTTGCTGCAAAAATA 1094
Db       1426 GTTAATAGTTTTATTATTTAGTAGTTAAAGTTATTTTATTTTTAATTGTTTAGTTAAAAG 1485

Qy       1095 AAAATTTTTCTGGTTGAAATAACTCAAGGCCTCTAGCGTTTTCCTTTATCTTAAAATACA 1154
Db       1486 AAGATAGTTATGTAAATTTTTATTTGTGGAAATTATTGTTGATAGTTAGTTTTTTATAAG 1545

Qy       1155 GGAAATAGCGATTGAAGTTAATTGACACTTAAG 1187
Db       1546 TTAAAAAAAAAATGGTATTTATTTAGCGTTTAG 1578
```

RESULT 11
US-10-239-676-224
; Sequence 224, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 224
; LENGTH: 6306
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-224

```
  Query Match             1.9%;  Score 68;  DB 15;  Length 6306;
  Best Local Similarity   49.8%;  Pred. No. 5.6e-05;
  Matches  201;  Conservative    0;  Mismatches  200;  Indels    3;  Gaps    1;

Qy        586 TTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTAT 645
Db       1745 TTATTTTTTATTTCGTTATTTTTTATCGTTTTATTTTTTTTATTTTTTATTTTTTTTAT 1804

Qy        646 TGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAGG 705
Db       1805 TTTTTTTATTTTTTATTTTTTATTTCGTTATTTTTTACGGCGTTATTTTTTAATTTTTAT 1864

Qy        706 TAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTT 765
Db       1865 TTTTTTATTTTTTTTATTTTTTTATTTTTTTTTATTTTTTTATTTTTTTTTATTTTTTTATTT 1924

Qy        766 TATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAA 825
Db       1925 TTTTTATTTTTTATTTTTTTTATTTTTTTATTTTTTTATTTTTTTTTATTTTTTTTTATTTTTT 1984

Qy        826 CATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTT-- 883
Db       1985 TTTATTTTTTTATTTTTTTTATTTTTTTATTTTTTTATTTTTTTTATTTTTTTTTATTTTTTT 2044
```

```
Qy        884 -TATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTC 942
              | | ||  ||  |  ||   |    | | | |  |||| | | |||||||   |
Db       2045 ATTTTTTTTATTTTTTTATTTTTTTTATTTTTTTATTTTTTTTATTTTTTTATTTCGAGG 2104

Qy        943 TTTTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGT 986
              ||| || || ||| | |    |   | |||   || || | |||
Db       2105 TTTAAAGGAAGTAAGTTATTTTTTCGATCGGAAAATTAAGACGT 2148


RESULT 12
US-10-240-453-332
; Sequence 332, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
;  TITLE OF INVENTION:  Transcription
;  TITLE OF INVENTION:  by Means of Assessing the Methylation Status of Genes Associated
;  TITLE OF INVENTION:  With DNA TRanscription
;  FILE REFERENCE: 5013.1009
;  CURRENT APPLICATION NUMBER: US/10/240,453
;  CURRENT FILING DATE:  2002-10-02
;  PRIOR APPLICATION NUMBER: PCT/EP01/03973
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 332
;   LENGTH: 6306
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-332

  Query Match             1.9%;  Score 68;  DB 15;  Length 6306;
  Best Local Similarity   49.8%;  Pred. No. 5.6e-05;
  Matches  201;  Conservative    0;  Mismatches  200;  Indels    3;  Gaps    1;

Qy        586 TTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTAT 645
              |||||  | | |    |  || | |   | |||  | ||||  ||  ||  | |||
Db       1745 TTATTTTTTTATTTCGTTATTTTTTTATCGTTTTATTTTTTTTTATTTTTTATTTTTTTAT 1804

Qy        646 TGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAGG 705
              |  | | |  ||| ||  | |||     || | | |   | ||||    ||||||
Db       1805 TTTTTTTATTTTTTATTTTTTATTTCGTTATTTTTTACGGCGTTATTTTTTAATTTTTAT 1864

Qy        706 TAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTT 765
              |  ||| | | ||| ||| || | |    ||| | |    ||  ||| | ||  |||
Db       1865 TTTTTTATTTTTTTTATTTTTTTATTTTTTTTATTTTTTTATTTTTTTTATTTTTATTT 1924

Qy        766 TATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAA 825
              | | | | |||   ||||  | |   |    || |  |||  | || ||| |
Db       1925 TTTTTATTTTTTTATTTTTTTTATTTTTTTATTTTTTATTTTTTTTATTTTTTTATTTTT 1984

Qy        826 CATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTT-- 883
              | |||  | |  | ||| |  | ||  |  |  ||    | |   | |||||||
Db       1985 TTTATTTTTTTATTTTTTTTATTTTTTTATTTTTTATTTTTTTTATTTTTTTATTTTTTT 2044

Qy        884 -TATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTC 942
              | | ||  ||  |  ||   |    | | | |  |||| | | |||||||   |
Db       2045 ATTTTTTTTATTTTTTTATTTTTTTTATTTTTTTATTTTTTTTATTTTTTTATTTCGAGG 2104

Qy        943 TTTTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGT 986
              ||| || || ||| | |    |   | |||   || || | |||
Db       2105 TTTAAAGGAAGTAAGTTATTTTTTCGATCGGAAAATTAAGACGT 2148


RESULT 13
US-10-240-589C-58
; Sequence 58, Application US/10240589C
; Publication No. US20040076956A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  DNA repair
;  FILE REFERENCE: 5013.1008
;  CURRENT APPLICATION NUMBER: US/10/240,589C
;  CURRENT FILING DATE:  2003-09-02
;  PRIOR APPLICATION NUMBER: PCT/EP01/03972
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 148
; SEQ ID NO 58
;   LENGTH: 13131
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (10848, 11413, 11540, 11554)
;   OTHER INFORMATION: n is a or g or c or t
US-10-240-589C-58

  Query Match             1.9%;  Score 68;  DB 17;  Length 13131;
  Best Local Similarity   48.1%;  Pred. No. 8.9e-05;
  Matches  223;  Conservative    0;  Mismatches  240;  Indels    1;  Gaps    1;

Qy        543 ATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGCTGTTTATTGGATGAATATTC 602
              |||   ||     | | |||| |  |   ||  |       | |||  |    ||||
Db       3635 ATTTTTTTTGAGTAGTTTTAATTTTATTATTTTGTTTCGTTTTTTATAATATTTTATTT 3694

Qy        603 GAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTATTGGTATCAACTTTCATG 662
                || ||   | | |    |  | || ||       | |||   | |     ||
Db       3695 TTTATAATGTTAAATTTAAGATTTTTTTTTTTTAAATTTTTATAATTGTATTTTAATATT 3754

Qy        663 CTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAGGTAATTTGTCTATTTAAT 722
              | |||   | || | ||||  || | |   |  |  |  |||||   ||| | |
Db       3755 TTTTATTTTTTTAGTTTATTAAGATTATGGTAATTTATAGTATGAATTTTGTTATATTTT 3814

Qy        723 TATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTTTATGTTTGTTTCATTTT 782
              ||||  | | |    |||| |   | | |||  | |  || |  | |||  |||||
Db       3815 ATTTATTTTATTTTATTGTATTTTTGGTTTTAAGATGTTTGTATTTTTTTTTATATTTT 3874

Qy        783 TAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAACATCTTTGATGAAGTAT 842
              |  |    || |  || |   ||  | ||||| ||  | | |   |  | |   || |
Db       3875 TTTTTTATGTTATGTTATTTTTTTTTTTTTTTAATTGTTATAGTTTGAATATTTGTGTTT 3934

Qy        843 TTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTTATATTAAATCATAATAAA 902
              |||  | |||  |||  | ||  |   ||||  |  |||| | |  || || |   |
Db       3935 TTATTAAATTTATAA-GTATATGATATTTGTATTGTTTTTATTTTATAGATTAGTTGATT 3993

Qy        903 ATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTTAAAGAGGTAGGATGAC 962
              |  | |  ||| ||| | |  | |||||  ||| ||   |||||  |||||| |
Db       3994 GAGTTTTAGAAAGTTTAATTTTTTTAGTATTATATAGTTGTTAGTAAGAGTTAGGATAAA 4053

Qy        963 CTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAATT 1006
              | || || || | | | | | ||   ||    |   ||  |||
```

```
Db       4054 ATTAAAGTAGCGTGATTTTAGAGTTTGTTATTTTAATTAGTATT 4097


RESULT 14
US-10-221-714A-27
; Sequence 27, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 27
;   LENGTH: 8666
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-27

  Query Match             1.9%;  Score 66.6;  DB 13;  Length 8666;
  Best Local Similarity   44.6%;  Pred. No. 0.00015;
  Matches  261;  Conservative    0;  Mismatches  324;  Indels    0;  Gaps    0;

Qy       525 TAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGCTG 584
Db        30 TAATTAGGTGTTTTTATTGTGTTATGATTTTTATTTTTAGTTGGGTTTTTTAGTTTGGTT 89

Qy       585 TTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTA 644
Db        90 TTAAAGAATAAAATAAAAGGTTTTTTTGTTTGAGTTTTATTTAATTTATTTTTTTTATTT 149

Qy       645 TTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAG 704
Db       150 ATAAGAGGTATTTTAAATTTTTACGATTTTTAAATTTTTTAATCGATTATTTTATTTTTA 209

Qy       705 GTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTT 764
Db       210 TTTTGTAAATAAGTTAATTTTTTTATTTATAGAAGGAAGTGTTTTAATTTTTTTTTCGTT 269

Qy       765 TTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACA 824
Db       270 TTGATTTTTTTTTTTTTTTAAATTTATGTATTTTTTGTGATAAAATTTATAATTATCGTT 329

Qy       825 ACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTT 884
Db       330 GTATTTTGAGTTTTATTTTTTTTATTATTTTTGAGGGATTTTAAGTTTTTAAAAATATTTT 389

Qy       885 ATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTT 944
Db       390 ATTTTGTTTGTGTATTTAATTTTTTTTTTATTTTTTTTTAATTTTTTTTTTTTTTTATTT 449

Qy       945 TTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAA 1004
Db       450 GGTATTTGTTTTTTTAGGTATATGTGTGTTTAGGTTTTTTTTATTTTTTATTTGTTTTAT 509

Qy      1005 TTCACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATA 1064
Db       510 TTTATGGTATAGGGTTTTGAATTATTATAATTAAGTTATGAAAGAGTAGTTAACGTAGTG 569

Qy      1065 TTAGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTT 1109
Db       570 TTTTTATTTTTTTGTTATTTTATTATTTTAGTTTTTTTTTTGGTT 614


RESULT 15
US-10-311-455-369
; Sequence 369, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 369
;   LENGTH: 8666
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-369

  Query Match             1.9%;  Score 66.6;  DB 15;  Length 8666;
  Best Local Similarity   44.6%;  Pred. No. 0.00015;
  Matches  261;  Conservative    0;  Mismatches  324;  Indels    0;  Gaps    0;

Qy       525 TAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGCTG 584
Db        30 TAATTAGGTGTTTTTATTGTGTTATGATTTTTATTTTTAGTTGGGTTTTTTAGTTTGGTT 89

Qy       585 TTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTA 644
Db        90 TTAAAGAATAAAATAAAAGGTTTTTTTGTTTGAGTTTTATTTAATTTATTTTTTTTATTT 149

Qy       645 TTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAG 704
Db       150 ATAAGAGGTATTTTAAATTTTTACGATTTTTAAATTTTTTAATCGATTATTTTATTTTTA 209

Qy       705 GTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTT 764
Db       210 TTTTGTAAATAAGTTAATTTTTTTATTTATAGAAGGAAGTGTTTTAATTTTTTTTTCGTT 269

Qy       765 TTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACA 824
Db       270 TTGATTTTTTTTTTTTTTTAAATTTATGTATTTTTTGTGATAAAATTTATAATTATCGTT 329

Qy       825 ACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTT 884
Db       330 GTATTTTGAGTTTTATTTTTTTTATTATTTTTGAGGGATTTTAAGTTTTTAAAAATATTTT 389

Qy       885 ATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTT 944
Db       390 ATTTTGTTTGTGTATTTAATTTTTTTTTTATTTTTTTTTAATTTTTTTTTTTTTTTATTT 449

Qy       945 TTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAA 1004
Db       450 GGTATTTGTTTTTTTAGGTATATGTGTGTTTAGGTTTTTTTTATTTTTTATTTGTTTTAT 509
```

```
Qy       1005 TTCACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATA 1064
              || |    | ||       | |    |       |  ||||  |       | |   | |
Db        510 TTTATGGTATAGGGTTTTGAATTATTATAATTAAGTTATGAAAGAGTAGTTAACGTAGTG 569

Qy       1065 TTAGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTT 1109
              ||  ||  | ||  || |  |   |   ||     ||||| |||||
Db        570 TTTTTATTTTTTTGTTATTTTATTATTTTAGTTTTTTTTTTGGTT 614
```

Search completed: July 3, 2004, 02:41:59
Job time : 1507 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 2, 2004, 15:43:37 ; Search time 8676 Seconds
(without alignments)
12349.643 Million cell updates/sec

Title: US-10-780-347-1
Perfect score: 3588
Sequence: 1 gtcgacttattgcattgatg.....gatnaaattgagccgcttaa 3588

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	91	2.5	1101	29	CNS00EVL	AL069706 Drosophil
	2	82.6	2.3	1200	13	BX436510	BX436510 BX436510
c	3	82	2.3	1094	29	CNS012FZ	AL101513 Drosophil
	4	80.2	2.2	1200	13	BX437739	BX437739 BX437739
c	5	79.2	2.2	999	13	BX380865	BX380865 BX380865
	6	78.8	2.2	714	28	AQ050116	AQ050116 nbxb0003b
	7	77.4	2.2	1201	9	AL536104	AL536104 AL536104
	8	77.2	2.2	712	13	BX416727	BX416727 BX416727
c	9	76.6	2.1	1101	29	CNS0039G	AL063921 Drosophil
	10	76.6	2.1	1101	29	CNS00HU4	AL074046 Drosophil
c	11	76.6	2.1	1201	13	BX439779	BX439779 BX439779
	12	76.4	2.1	1201	13	BX462207	BX462207 BX462207
	13	76.2	2.1	1200	13	BX415878	BX415878 BX415878
c	14	75.8	2.1	1200	13	BX437739	BX437739 BX437739
	15	75.8	2.1	1201	13	BX366070	BX366070 BX366070
	16	75.6	2.1	1128	13	BX337294	BX337294 BX337294
c	17	75.2	2.1	527	14	CA726232	CA726232 wetls.pk0
c	18	75.2	2.1	781	29	CNS009DO	AL053444 Drosophil
c	19	75	2.1	1201	9	AL565455	AL565455 AL565455
	20	74.8	2.1	1268	28	BZ577630	BZ577630 msh2_550.
c	21	74.6	2.1	1195	13	BX355698	BX355698 BX355698
	22	74.4	2.1	1101	29	CNS0039G	AL063921 Drosophil
	23	74.2	2.1	576	29	CNS035N7	AL228940 Tetraodon
	24	73.8	2.1	1098	13	BX377526	BX377526 BX377526
c	25	73.8	2.1	1101	29	CNS012I9	AL101595 Drosophil
	26	73.6	2.1	1201	13	BX406288	BX406288 BX406288
c	27	73.6	2.1	1201	13	BX446437	BX446437 BX446437
c	28	73.4	2.0	996	29	CNS00FUH	AL071063 Drosophil
c	29	73.2	2.0	870	29	CNS009XJ	AL054462 Drosophil
c	30	73.2	2.0	1175	28	BZ696793	BZ696793 SP_Ba008
c	31	73.2	2.0	1201	13	BX461824	BX461824 BX461824
	32	73	2.0	1201	13	BX335216	BX335216 BX335216
	33	72.8	2.0	991	29	CNS00I2D	AL074343 Drosophil
	34	72.8	2.0	997	13	BX419578	BX419578 BX419578
	35	72.8	2.0	1084	29	CNS06PLB	AL409509 T7 end of
c	36	72.8	2.0	1200	13	BX437758	BX437758 BX437758
	37	72.8	2.0	1200	29	CNS016EL	AL106647 Drosophil
c	38	72.6	2.0	614	29	CNS0152H	AL104915 Drosophil
	39	72.6	2.0	881	29	CG767277	CG767277 TcB43.2_A
	40	72.6	2.0	1201	13	BX445758	BX445758 BX445758
	41	72.6	2.0	1348	29	CG749499	CG749499 P043-4-A0
c	42	72.4	2.0	706	29	CE784276	CE784276 tigr-gss-
c	43	72.4	2.0	1061	13	BX437039	BX437039 BX437039
	44	72.4	2.0	1139	28	AQ897537	AQ897537 HS_3153_A
c	45	72.4	2.0	1165	13	BX338369	BX338369 BX338369

ALIGNMENTS

RESULT 1
CNS00EVL/c
LOCUS CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_lib="RPCI-98"
/note="end : T7"

ORIGIN

Query Match 2.5%; Score 91; DB 29; Length 1101;
Best Local Similarity 34.9%; Pred. No. 1.8e-09;
Matches 217; Conservative 106; Mismatches 296; Indels 2; Gaps 1;

Qy 585 TTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTA 644
Db 1097 TATWTWTTTWTTTTTWWTATWTTAWATATATAAWTAATTWWTWTATATWTAWWWTATWAT 1038

Qy 645 TTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAG 704
Db 1037 ATATTAWTTTWTAAAATATAWWTTAATAAATATATWTAWTAWWATATATAWATAWAAWTA 978

Qy 705 GTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTT 764
Db 977 TAWAWTWWAATWAATWWATAWAWTTTAATAAWAATATWAWWWTWTTATTTTTTTTTTTTT 918

Qy 765 TTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACA 824
Db 917 WTTTWTATWTWTTTATATWWTWWAAAAWAWWAAWTWTATTTTTTWTTATWATTWATATTA 858

Qy 825 ACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTTT 884
Db 857 WATTTTTTWTWTATTWWTATATWATWWTATWTWTWTTTWATWTATWTATTTATWTWTWTT 798

Qy 885 ATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTT 944
Db 797 WTATTWTATWTATAWWWTWWATWWWTAWAAWTATATATWWATATTATWWTWATTWTTWTT 738

Qy 945 TTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAA 1004
Db 737 TTATTWTTTTAAWTTATWTATWATTWTTWATWTTTTTTTTTTTTTTTTWTATTTWAAATTA 678

Qy 1005 TTCACCTCAAAGATACACTATTGGCAAATTGA--CAAATATGTCACTTCGTATGAAACAA 1062
Db 677 TTTAATTAWWWAATWAWTTATTTAATAAWTAATTAAATTATTAAAAATTTTATWAAAAAA 618

Qy 1063 TATTAGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTTGAAATAACTCAAG 1122
Db 617 TATTTTTTTATATTAATWAWATTTATAWATTAATATAWWTTWTAAATTWTTAATWTATTW 558

Qy 1123 GCCTCTAGCGTTTTCCTTTATCTTAAAATACAGGAAATAGCGATTGAAGTTAATTGACAC 1182
Db 557 WTTTAAATTATWTTTTTTTTTAWTAWAWWTWAAAWAATWAWWWWWAAAAATWWWWTWTTT 498

Qy 1183 TTAAGCAAATAGTCAACCTAA 1203
Db 497 TTWATTWATAAWTTTWTTTWW 477

RESULT 2
BX436510
LOCUS BX436510 1200 bp mRNA linear EST 15-MAY-2003
DEFINITION BX436510 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP002YK09 5-PRIME, mRNA sequence.
ACCESSION BX436510
VERSION BX436510.1 GI:30770190
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1200)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 5393.f For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0CAP002AF05QP1&cluster=5393.f. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0CAP002AF05QP1.

FEATURES Location/Qualifiers
source 1. .1200
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP002YK09"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/note="Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 2.3%; Score 82.6; DB 13; Length 1200;
Best Local Similarity 41.2%; Pred. No. 1.3e-07;
Matches 175; Conservative 47; Mismatches 203; Indels 0; Gaps 0;

Qy 543 ATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGCTGTTTATTGGATGAATATTC 602
Db 726 ATTTATTWWTTTTTWWAWTWAAATAAAAWTATTTWWTAWATTTTTTWTAAAAATWAAAWA 785

Qy 603 GAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTATTGGTATCAACTTTCATG 662
Db 786 TAATWNTTTWTWWWTTTATATTTTTTTTTTTWWTTTWTTTTTTWATTWAAAATTTTTTTA 845

Qy 663 CTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAGGTAATTTGTCTATTTAAT 722
Db 846 ATATTTTTTTTAATWAAWTTTTAATTATTWAAAAWWTTTTWTTTTTTTWWTWTTTTTTMN 905

Qy 723 TATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTTTATGTTTGTTTCATTTT 782
Db 906 TTTTTTTTTTTTTTTTWWWTTTTTTTTTTTTAAAAATTTTATTATTTTTTTTWAAATAT 965

Qy 783 TAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAACATCTTTGATGAAGTAT 842
Db 966 TWTTWTTTTATWWAAAWTTTTTTATAWTTTTTTAWTWTTTTAAWWATATTAATTTTTTTT 1025

Qy 843 TTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTTATATTAAATCATAATAAA 902
Db 1026 WWATTATATTTTATAAAWAATAATAWTTTTAAAATTTTTAATTWAATAAAAWTTWTTAWA 1085

Qy 903 ATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTTAAAGAGGTAGGATGAC 962
Db 1086 TTWAWTWTTATATTTAWAWTTAWWTAATTTTAAWTTWTTTTTAAAAAWAAAAAAAAAAAA 1145

Qy 963 CTAAA 967
Db 1146 AAAAA 1150

```
RESULT 3
CNS012FZ/c
LOCUS       CNS012FZ                1094 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN07B02 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL101513
VERSION     AL101513.1  GI:5613124
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1094)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1094
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN07B02"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             2.3%;  Score 82;  DB 29;  Length 1094;
  Best Local Similarity   36.4%;  Pred. No. 1.8e-07;
  Matches  192;  Conservative   88;  Mismatches  246;  Indels    2;  Gaps    1;

Qy        558 TATTAAATAAACAGTTCTAAAAGGCTGTTTATTGGATGAATATTCGAAATTATCACATAA 617
              :|  |||    |    : : ::   | |||:::   || | |   |||| | | :  |
Db       1090 WAAAAAAGATKGAARAWKGRTWDTTTTTTTDWWWTTTGGAAAARAGAAAATTTTWTTTTT 1031

Qy        618 TAATTGATGCTATTATTACTTGCTGTATTGGTATCAACTTTCATGCTCTATACATGTAAT 677
              | ::|  |  :::| ||   :  | | || :::  |  |||| :      ||:|    |  |
Db       1030 TGRWTTKTTAKRWTTTTTKAWTTTTTTTTADWRRAATTTTTWDRWTAATAWATTAWTTTT 971

Qy        678 ATATTTCGAGTTAGACCTTAATTCAAGGTAATTTGTCTATTTAATTATTATCTGAATAAT 737
              |:|:|  ||    :  ||:| | :|   |::::   :|  |:|  :  :| : :  | |
Db        970 TTWTWTGAAGRRGARAATTWAATTWAAAAAWWWWWATWAAATWAAADAGRTKWAWTTTAG 911

Qy        738 ATGTAATCGATTGCTTTGTGG--TTATTTTTATGTTTGTTTCATTTTTAATGACGGTGAG 795
              :| :  |:||  ||   ||  ::|  :|| |   | ||| |||:| || ||   | :
Db        910 GRRTGRGTGRTTWTTTAARGGRTWWAAAWTTTTWGATTTTTWATTWTGAAAGAWTTTAWW 851

Qy        796 CTTGTGCATTCATATTTTTTATGATGACAACATCTTTGATGAAGTATTTAAGATATTGTT 855
              || |   ||  || |||:  | |  : ||  | | |  :: ::|||   : :|| ||
Db        850 TTTTTWTTTTTTTAATTTWATTTAAAWTAAATTTTATTAAARWAWWTTTTTAWADTTTTT 791

Qy        856 AATGCATGAGGGGTTTGCGTGTATTTTTTATATTAAATCATAATAAAATCAACAATATAT 915
              |     |:::::|| : '|  |:||    |   :| :|  |  :|: || :|||| |
Db        790 TTTTTTRGRRRRRTTGRWWTTAAWTTRAAAAGGTTTWTRWTTTTTTWAWAAAARATATTT 731

Qy        916 GTTATTTTGTGTCTTTTTATAGTGTTCTTTTAAAGAGGTAGGATGACCTAAAGGTCGCCT 975
              || |:|| : : |||||:|  :    | :||:   | :|     |::|  |    |
Db        730 TTTTTWTTWWWWTTTTTTRTTTWAAAAAATARAARTTTTTDGGGTGRTTWRATATAAATT 671

Qy        976 AAATATGGCGTAAATTGCCATTGCTATAATTCACCTCAAAGATACACTATTGGCAAATTG 1035
```

```
              :||: |   | ::| |    ||:  | :::| |  |  :||:|    | | |: ||||:|
Db        670 WAAKTTTTTGGRWARTTTTTTTRRAAAWWWTGAWATGGRAGVTTTGRTTTGGRKAAATKG 611

Qy       1036 ACAAATATGTCACTTCGTATGAAACAATATTAGTAGATGTTGTTTTTG 1083
                || | |      | ||  ::: ||  |  || || ||||||
Db        610 TWAATTTTTGGTTTGGAATTTAAGWRRWTTTTTTTAATTTTTTTTTG 563


RESULT 4
BX437739
LOCUS       BX437739                1200 bp    mRNA    linear   EST 22-MAY-2003
DEFINITION  BX437739 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP008YG24
            3-PRIME, mRNA sequence.
ACCESSION   BX437739
VERSION     BX437739.1  GI:31018315
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1200)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 534.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0CAP008BD12NP1&cluster=534.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0CAP008BD12NP1.
FEATURES             Location/Qualifiers
     source          1. .1200
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0CAP008YG24"
                     /tissue_type="THYMUS"
                     /clone_lib="Homo sapiens THYMUS"
                     /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             2.2%;  Score 80.2;  DB 13;  Length 1200;
  Best Local Similarity   35.3%;  Pred. No. 4.5e-07;
  Matches  270;  Conservative  112;  Mismatches  379;  Indels    4;  Gaps    2;

Qy        515 TTTGGTGCTTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTC 574
              |||  |  ||| |||  : | |  |  |     |||| ||   : ||  :||      ||
Db        416 TTTATTTTTTTTTATTTAWNATTATTTTTTTAATTATTATTTATTWTTTTTWTATTTTTTTT 475

Qy        575 TAAAAGGCTGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATT 634
              | |    | :|| ||  :    |||      || |     | | ||| ||  | || ||
Db        476 TTATTAATTTWTTTTTAAWTATTGATTATTTTTTTTTTTTTTTTTATTAATAATTTTTTT 535

Qy        635 ACTTGCTGTATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACC 694
              ||  |:| |    |  |:|  |  | |||    :  | ||| :|||     |||
Db        536 TATTATTKTTTNAAATTTKTTTWTATTTATATTATTATDTTTATAAWTTTTTTTTTATTTT 595

Qy        695 TTAATTCAAGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTT 754
              |  ||      | |||| | |  ||  || |: | | :|| |: | |      ||  |:|
Db        596 ATTTTTTTTTWTTATTTTTATTATTTTTTTTKTTTTTWATTAWTTTTTTAATTTTTATKT 655

Qy        755 GTGGTTATTTTTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTT 814
              |: |    |      ||:||| ::   : |     :| |   |:|  |::  ||||||
```

```
Db        656 TTKTTATAATAATAWATTRTTTWWWAAAWTAATTTTRTTATTATKTTTAWWATTATTTTA 715

Qy        815 TATGATGACAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCG 874

Db        716 NTTTATTGTTTTTWTTGAAATAAAAAATAATTTDAAKTTTTATATATTATWTTTTTATGA 775

Qy        875 TGTATTTTTTATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTA 934

Db        776 TATAWTTTTAAWAAWAWKTTWTTATAAAATAAAAATTAATTAATAWTTTATATWWAATWA 835

Qy        935 TAGTGTTCTTTTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCC 994

Db        836 WWAWTAAATTATTTAAWTAAATDAAAAWAWAWTAAAATAAAAAAAAATATAAAAAATTTA 895

Qy        995 ATTGCTATAATTCACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTA 1054

Db        896 AAATTGATAATATTATTTATATATA---TATWTWATWAATTATAAAAAAATTNTATAAAA 952

Qy       1055 TGAAACAATATTAGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGGTTGAAAT 1114

Db        953 AAWATTAAKKAKDTKAKDAAAAAKTKTKTDAAAAAAAAADAAKDKKKGTATKAWAAKATK 1012

Qy       1115 AACTCAAGGCCTCTAGCGTTTTCCTTTATCTTAAAATACAGGAAATAGCGATTGAAGTTA 1174

Db       1013 AAKKKKAKGTADAAADATADKKKAAKAAAAKKAAAAKA-AKAAAKKKKTKRTKKKAKAKK 1071

Qy       1175 ATTGACACTTAAGCAAATAGTCAACCTAACAGAGCAGGAACCTATGCCTTTGTCAAAGCA 1234

Db       1072 AKAAAAAKDADKKAAAKAADAKAAAAAWGAAKAKAAKKAKAADAKKKKDKAAAAAAAKKK 1131

Qy       1235 TCAAATTGAGCAACTTTCTAAACCTCTGAGTGATGATTCGATCTG 1279

Db       1132 KKAWKKAAKKKAAAAAAAWKKKDAADWDADAKAKDAKKDDATAKD 1176
```

RESULT 5
BX380865/c

```
LOCUS       BX380865               999 bp    mRNA    linear   EST 08-MAY-2003
DEFINITION  BX380865 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI058YE10 3-PRIME, mRNA sequence.
ACCESSION   BX380865
VERSION     BX380865.1  GI:30439622
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 999)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 3221.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI058BC05NP1&cluster=3221.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI058BC05NP1.
FEATURES             Location/Qualifiers
     source          1. .999
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI058YE10"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
```

```
  Query Match             2.2%;  Score 79.2;  DB 13;  Length 999;
  Best Local Similarity   32.3%;  Pred. No. 7.5e-07;
  Matches  145;  Conservative  109;  Mismatches  191;  Indels    4;  Gaps    1;

Qy        515 TTTGGTGCTTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTC 574

Db        934 WWTTTWTTTNWWAWWWTTTTTTWTTAWTTWWWTTTTWTTATAAAWAWAAAAAAAAW 875

Qy        575 TAAAAGGCTGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATT 634

Db        874 AAAAAAAAATTTTTTWWTAWTTTTNTTNAAWWTTWTTWMTTNTTTWMAHHTTHTMTTTHT 815

Qy        635 ACTTGCTGTATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACC 694

Db        814 AHAANTATTWTWAATWWWAWATTTTAT----TAMMHNTTTTTTHHTTTWTHHMNTTTTHT 759

Qy        695 TTAATTCAAGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTT 754

Db        758 TTTTHTTTTTTWTTTTTTWTWTTTTTMTMMTTNNTTTHHTTMHMTMHMMMMMMTTTTHT 699

Qy        755 GTGGTTATTTTTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTT 814

Db        698 MTTTTTTTTTTYYTTTTTTTMMNMHTTTMNTMMYMMMMTHMCTTHMMMWTHTWTMHTHTT 639

Qy        815 TATGATGACAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCG 874

Db        638 TTTTMMMMHMMMMMMTTTTHTHTTTTHTHTTTMTTMMHNMMTTAAWAWNAAAAWTTWWW 579

Qy        875 TGTATTTTTTATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTA 934

Db        578 WTTWTWTWWWWWAWAAANAAAAAAATAAAWTAAAAATAAAYTMYCHNTTTTTHTMMHTTM 519

Qy        935 TAGTGTTCTTTTAAAGAGGTAGGATGACC 963

Db        518 TAATAWWHVTTCVVVAVNVTNAAANNACV 490
```

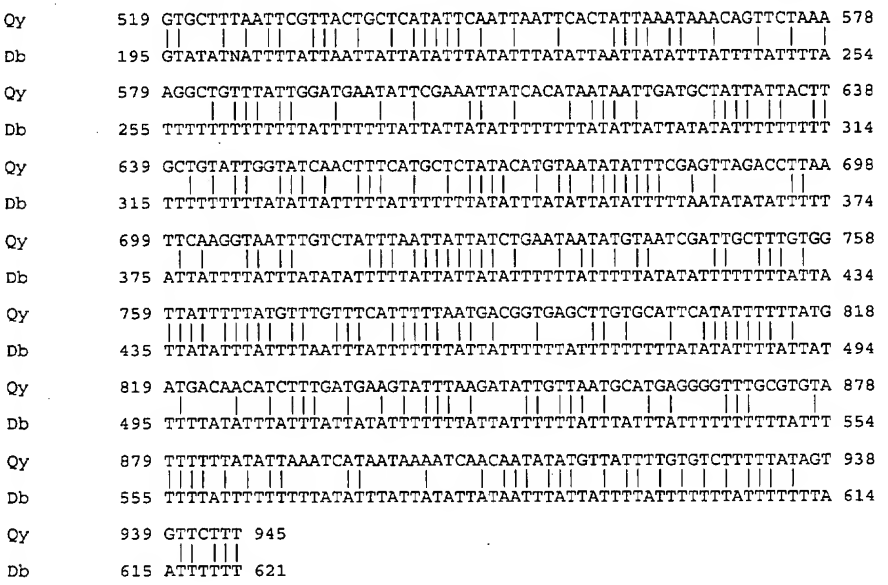
RESULT 6
AQ050116

```
LOCUS       AQ050116               714 bp    DNA     linear   GSS 24-MAR-1999
DEFINITION  nbxb0003bDO1r CUGI Rice BAC Library Oryza sativa (japonica
            cultivar-group) genomic clone nbxb0003G00r, genomic survey
            sequence.
ACCESSION   AQ050116
VERSION     AQ050116.1  GI:3324505
KEYWORDS    GSS.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 714)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GGAAACAGCTATGACCATG
            Class: BAC ends
            High quality sequence stop: 143.
FEATURES             Location/Qualifiers
     source          1. .714
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
```

/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="nbxb0003G00r"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/clone_lib="CUGI Rice BAC Library"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

ORIGIN

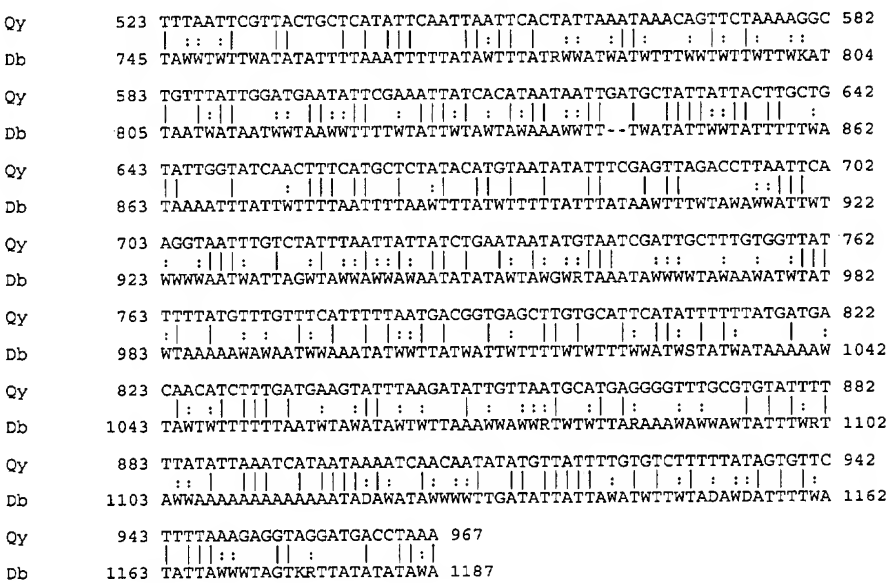
Query Match 2.2%; Score 78.8; DB 28; Length 714;
Best Local Similarity 48.9%; Pred. No. 9.4e-07;
Matches 209; Conservative 0; Mismatches 218; Indels 0; Gaps 0;



RESULT 7
AL536104
LOCUS AL536104 1201 bp mRNA linear EST 31-MAY-2003
DEFINITION AL536104 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone CS0DF022YC18 5-PRIME, mRNA sequence.
ACCESSION AL536104
VERSION AL536104.2 GI:31260974
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 13, 2001 this sequence version replaced gi:12799597.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DF022BB09QP1.
FEATURES Location/Qualifiers
source 1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF022YC18"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 2.2%; Score 77.4; DB 9; Length 1201;
Best Local Similarity 34.4%; Pred. No. 1.9e-06;
Matches 153; Conservative 93; Mismatches 197; Indels 2; Gaps 1;



RESULT 8
BX416727
LOCUS BX416727 712 bp mRNA linear EST 15-MAY-2003

```
DEFINITION  BX416727 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
            CS0DA011YI14 5-PRIME, mRNA sequence.
ACCESSION   BX416727
VERSION     BX416727.1  GI:30765629
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 712)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DA011BE07QP1.
FEATURES             Location/Qualifiers
     source          1. .712
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DA011YI14"
                     /tissue_type="NEUROBLASTOMA"
                     /clone_lib="Homo sapiens NEUROBLASTOMA"
                     /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN
```

```
  Query Match             2.2%;  Score 77.2;  DB 13;  Length 712;
  Best Local Similarity   21.7%;  Pred. No. 2.1e-06;
  Matches   91;  Conservative  153;  Mismatches  176;  Indels    0;  Gaps    0;

Qy        541 ATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGCTGTTTATTGGATGAATAT 600
              | | : ||  || : |  |  ||| ||| |     ||||        |   | || |:
Db        240 AAAAWAAAAAAAAAWAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAWAAAAW 299

Qy        601 TCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTATTGGTATCAACTTTCA 660
                :||::|: : ::|:::|:: :|  :||:::|: ||  | | ||  | |     |||
Db        300 AWAWAAWWAWWWAWWAWWWAWWWTTWWATWWWTWWTTTTTTTTTTTTTTTTTTTTWT 359

Qy        661 TGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAGGTAATTTGTCTATTTA 720
              |  | | |    | |  | | |||     ||       ||  ||     :|  :|| | | |||
Db        360 TTTTTTTTTTTTTTTTTTTTTTTTWTTTTTTTTTTTTTTTTTKTKTKTTTTTTTTTT 419

Qy        721 ATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTTTATGTTTGTTTCATT 780
              || || : |    |  | ::|  | : ::  |||:|: :: ::|:| |:::||::   ::
Db        420 TTTTTTTKTTTTKTKTTKKKTKKTKKKKKTKTTTKTKTKKKKKTKTKTKKKTGKKGKKKK 479

Qy        781 TTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAACATCTTTGATGAAGT 840
              |::  ::  :|:| : :::::  ::  : ::::::: :: ::       : :::: ::  ::
Db        480 TKKKKKKKKKGKGKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKDKKKKKKKKKK 539

Qy        841 ATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTTATATTAAATCATAATA 900
                :::  : : :::::  :: |:: :::::::: :::: :::::: : ::   :  :  :
Db        540 KKKKKKKKKKKKKKKKKKKKAKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 599

Qy        901 AAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTTAAAGAGGTAGGATG 960
                :     : : :::: :::::::: ::::: : ::::: ::::   : ::: :: ::
Db        600 KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 659
```

```
RESULT 9
CNS0039G/c
LOCUS       CNS0039G                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN
```

```
  Query Match             2.1%;  Score 76.6;  DB 29;  Length 1101;
  Best Local Similarity   17.5%;  Pred. No. 2.8e-06;
  Matches  124;  Conservative  299;  Mismatches  282;  Indels    2;  Gaps    1;

Qy        614 ATAATAATTGATGCTATTATTACTTGCTGTATTGGTATCAACTTTCATGCTCTATACATG 673
              ::::|: :: ::  ::|:::: ::: :  : ::  ::  | |:: ||  : ::::  ::
Db       1091 WDRDTRKDDWDWTKWWTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWWWWWWWWATWD 1032

Qy        674 TAATATATTTCGAGTTAGACCTTAATTCAAGGTAATTTGTCTATTTAATTATTATCTGAA 733
              |::: ::::   |:|:  :  :  |:  :: | ::   :   :  |:|: |  ::  : :
Db       1031 TWWDKWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADGAGRRDGGRKR 972

Qy        734 TAATATGTAATCGATTGCTTTGTGGTTATTTTTATGTTTGTTTCATTTTTAATGACGGTG 793
              :: :: :: :: :  : ::  :  ::||:::::::::::::::   :: :|::::  |::
Db        971 KDKKDRKDGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKADDDDGAGDK 912

Qy        794 AGCTTGTGCATTCATATTTTTTATGATGACAACATCTTTGATGAAGTATTTAAGATATTG 853
              ::  :|:: ::: :  :::::::: :: : |   : : | |: | ::::: :|:|:::|
Db        911 DDDGKGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWADADWWTW 852

Qy        854 TTAATGCATGAGGGGTTTGCGTGTATT--TTTTATATTAAATCATAATAAAATCAACAAT 911
              : ||:: : ::: : : :: :: :|:   | : ::: ::| :   |:::::: |:  ::
Db        851 DAAADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRKRADDKRD 792

Qy        912 ATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTTAAAGAGGTAGGATGACCTAAAGGTC 971
              | :::::| |:| : | || :|::::::|: :| :||::  : :::::::     | :::
Db        791 AADDRDDAATWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDRAGTAGRKWR 732

Qy        972 GCCTAAATATGGCGTAAATTGCCATTGCTATAATTCACCTCAAAGATACACTATTGGCAA 1031
              :  :::::: :   ::|: :: :::  : ::    |       :  : :  ||::: :|
Db        731 RTWKRRWKRRDTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKRRRRDRATWDRTDA 672
```

```
Qy       1032 ATTGACAAATATGTCACTTCGTATGAAACAATATTAGTAGATGTTGTTTTTGCTGCAAAA 1091
Db        671 WWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDWAKWDWKTRAD 612

Qy       1092 ATAAAAATTTTTCTGGTTGAAATAACTCAAGGCCTCTAGCGTTTTCCTTTATCTTAAAAT 1151
Db        611 RWDRWAADTWTDARKADRDWAKARAWRARRDRARAARADRRWTTKGKTTTATWTTWAARA 552

Qy       1152 ACAGGAAATAGCGATTGAAGTTAATTGACACTTAAGCAAATAGTCAACCTAACAGAGCAG 1211
Db        551 AWWAWWAWATTTATWTTTWTTTWTTTTTTTTTTTWTTTAAWWWAAWWTATWAAWTAAAWA 492

Qy       1212 GAACCTATGCCTTTGTCAAAGCATCAAATTGAGCAACTTTCTAAACCTCTGAGTGATGAT 1271
Db        491 AAAAAAWAATTTTTTTTTTTTWAWTAAAWWTAWTWTTTWTTTTWWAATTTTTWTWTTW 432

Qy       1272 TCGATCTGTGGCGTTTATCTTAAACTGGAAAAAAGTGCTTTTCGCCC 1318
Db        431 WAATTATTTTWTTTWTAAAAAAAAWWWWWTWTATATKCCCCCCYCCC 385
```

RESULT 10
CNS00HU4
LOCUS CNS00HU4 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR35O18 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL074046
VERSION AL074046.1 GI:4953725
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR35O18"
/clone_lib="RPCI-98"
/note="end : T7"
ORIGIN

Query Match 2.1%; Score 76.6; DB 29; Length 1101;
Best Local Similarity 34.7%; Pred. No. 2.8e-06;
Matches 191; Conservative 11; Mismatches 348; Indels 0; Gaps 0;

```
Qy        628 TATTATTACTTGCTGTATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAG 687
Db         48 TTTTTTTTTTTTKTTTTTTNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 107

Qy        688 TTAGACCTTAATTCAAGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGA 747
Db        108 TTTGTDTDGATGTNATAAAAAAAATTTAAWATTATTTATATATTCHDWTTTNTTTATTTN 167

Qy        748 TTGCTTTGTGGTTATTTTTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCA 807
Db        168 TTTTTNTTTTNTTATTTTTTTTTTTTTTTATTTTTWNTTTNTTTTTTTTTTTTATATTTT 227

Qy        808 TATTTTTTATGATGACAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGG 867
Db        228 TTTTTTTTTNTTTNAATTATTTATTAASAKGAGAGAGGNAAGAATTAGAGGGVGGGAGTN 287

Qy        868 GTTTGCGTGTATTTTTTATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGT 927
Db        288 NNTNNANTNNNNTNNNNANATTNTTTNNTAANNNNTNANNNNTNNTANNTNNNTTTTTTT 347

Qy        928 CTTTTTATAGTGTTCTTTTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTA 987
Db        348 TTTWTTATTTTNTTNTTTTTTWTTTTTTTANGNNNANANNTTTGGNNANNNCNGNNNNAN 407

Qy        988 AATTGCCATTGCTATAATTCACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCA 1047
Db        408 NNATTTAATTTTTTTTTATWAAATANATTTTTAATTATWGTNNNNNNNAANTNANTANTN 467

Qy       1048 CTTCGTATGAAACAATATTAGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTTCTGG 1107
Db        468 NTATANNTNNNTNATTACANNANNAAANNNNANNTNNNNNNAAAANNAAAANNTNNNTNN 527

Qy       1108 TTGAAATAACTCAAGGCCTCTAGCGTTTTCCTTTATCTTAAAATACAGGAAATAGCGATT 1167
Db        528 TNTAANNNNNNNANNGNNNGGGNNNANNNNNNNNANTNWTNAATAGNNNNNNNATTTTTT 587

Qy       1168 GAAGTTAATT 1177
Db        588 TTTNTTTNTT 597
```

RESULT 11
BX439779/c
LOCUS BX439779 1201 bp mRNA linear EST 15-MAY-2003
DEFINITION BX439779 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE014YF05 3-PRIME, mRNA sequence.
ACCESSION BX439779
VERSION BX439779.1 GI:30771778
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 3370.r For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DE014CC03NP1&cluster=3370.r. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DE014CC03NP1.
FEATURES Location/Qualifiers
source 1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE014YF05"
/tissue_type="PLACENTA"

/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 2.1%; Score 76.6; DB 13; Length 1201;
Best Local Similarity 39.1%; Pred. No. 2.8e-06;
Matches 160; Conservative 64; Mismatches 184; Indels 1; Gaps 1;

```
Qy       515 TTTGGTGCTTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTC 574
Db      1114 TTTTTTTWTTTWAATTTAWTWTTTWTTTTTTTTTTTTWAWTTTTWTTWAWTTWWATAAAW 1055

Qy       575 TAAAAGGCTGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATT 634
Db      1054 AAAWAAAATWWWTTTTWAATAWATWTTTWAATTTTTAWAWTWATTTTTTTTTWTTTTTTT 995

Qy       635 ACTTGCTGTATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACC 694
Db       994 TTTWTAATTTWWAWTTTTTAATWTTTTTATAWWTAWATWWATTWTTTTTATWAWTTTAAA 935

Qy       695 TTAATTCAAGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTT 754
Db       934 AWTWTWAAWTTTAWTATWWTAAAAWATTWTWTWTAAAWAWATTWAAWTATTWWAWWWTTT 875

Qy       755 GTGGTTATTTTTATGTTTGTTTCATTTT-TAATGACGGTGAGCTTGTGCATTCATATTTT 813
Db       874 TAAAAWTTWTTTTTTWTWWTTTAWTTTTAAAAWTWAAWTTATWTATWAAAAAAWTATKTT 815

Qy       814 TTATGATGACAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGC 873
Db       814 ATAAATTTTWTAAATTATATTTAAAAWAATTAAATTTTWAAATWTAMATTTATATTTTAW 755

Qy       874 GTGTATTTTTTATATTAAATCATAATAAAATCAACAATATATGTTATTT 922
Db       754 TTAWATTTTTATTDTAWAATTTTTAWTTWATTTTAATAATTTWTTTTWT 706
```

RESULT 12
BX462207
LOCUS BX462207 1201 bp mRNA linear EST 22-MAY-2003
DEFINITION BX462207 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA clone CS0DG004YB03 5-PRIME, mRNA sequence.
ACCESSION BX462207
VERSION BX462207.1 GI:31023422
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 24.r For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DG004CA02QP1&cluster=24.r. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DG004CA02QP1.
FEATURES Location/Qualifiers
source 1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

/clone="CS0DG004YB03"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
/note="Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 2.1%; Score 76.4; DB 13; Length 1201;
Best Local Similarity 42.0%; Pred. No. 3.1e-06;
Matches 182; Conservative 37; Mismatches 214; Indels 0; Gaps 0;

```
Qy       515 TTTGGTGCTTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTC 574
Db       595 TTTTTTTTTTTTTTTTTTTWATTTTTTTWTTTTWAWTTTTTTTTTWAWAAAAATTTT 654

Qy       575 TAAAAGGCTGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATT 634
Db       655 TAWWAAWWWTWTTTTWTTTTWWAAAAAAAAAAAAWTAWTTTTTTTTTTWWWAWTTTTWTW 714

Qy       635 ACTTGCTGTATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACC 694
Db       715 TTTTAAAAAWTTTAWTTTTTTWTTTTTTTTTTAAAAAWWTTTTTTTTTTTTWTTTTWWAAA 774

Qy       695 TTAATTCAAGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTT 754
Db       775 ATTTTTTTAAAAATTTTTTTTTTTTTTTTTWTTRTTATTTTTTTTTTTTTTTTTTTTTT 834

Qy       755 GTGGTTATTTTTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTT 814
Db       835 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTWTAWTTTWWTTATTTATTTATTTTTTTTT 894

Qy       815 TATGATGACAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCG 874
Db       895 TARAATTTTWAAATTTTTTTAAATWAATATTTTTTTTTTTTTTTTTTTTTTTTTTTTT 954

Qy       875 TGTATTTTTTATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTA 934
Db       955 TTTTTRWAAATWTTTTATTTTTTTTTTTWTTWWAAANATTTTTTTTTTRAAWKTTTTTWT 1014

Qy       935 TAGTGTTCTTTTA 947
Db      1015 TAATTTATTTAKA 1027
```

RESULT 13
BX415878
LOCUS BX415878 1200 bp mRNA linear EST 15-MAY-2003
DEFINITION BX415878 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP008YI04 5-PRIME, mRNA sequence.
ACCESSION BX415878
VERSION BX415878.1 GI:30765550
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1200)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0CAP008BE02QP1.
FEATURES Location/Qualifiers

```
     source          1. .1200
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0CAP008YI04"
                     /tissue_type="THYMUS"
                     /clone_lib="Homo sapiens THYMUS"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             2.1%;  Score 76.2;  DB 13;  Length 1200;
  Best Local Similarity   36.0%;  Pred. No. 3.5e-06;
  Matches  247;  Conservative   86;  Mismatches  352;  Indels    2;  Gaps    1;

Qy        523 TTTAATTCGTTACTGCTCATATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGC 582
              || | ||    | |  :   :::  : :| ||    |||:||   |    | |
Db        497 TTAATTTTAAAAWTTWWATAWWWTTWAWATTTTTTAATTWAAATTAATWATTTTTTTTTA 556

Qy        583 TGTTTATTGGATGAATATTCGAAATTATCACATAATAATTGATGCTATTATTACTTGCTG 642
              | | ||  |  |: | |  :  |: |   :|:| |::|  |  : || |:  ||  |
Db        557 AATATTTTAAAAAAWAAATWAWTTWTTTTTWTWAAAWWTTTTATWTTTTTWTWTTWTTT 616

Qy        643 TATTGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCA 702
              ||||  ||| |  |:|  |  |  ||: || |||    |   :: :: | :  ||:
Db        617 TATTTTTATAATWTWTTTTAATAAATWAATATAAATATTAAWTWRAWWAAWTWATATWTT 676

Qy        703 AGGTAATTTGTCTATTTAATTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTAT 762
              |||||  |  || |  ||   ||  |  || :||| :|||  : :  |||       |
Db        677 TTTTAATTAGATTAATATATATATAATTTTATTWTATTWAATWWWAWTTTTTWAYAAGAA 736

Qy        763 TTTTATGTTTGTTTCATTTTTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGA 822
              :|  |  |   |  || | ||  | :   |   | |  :|: |:    | | | |: |
Db        737 AWTATTAATACATANATATATATATAWDCATAAAATATWAWTWTAWTAAATATTTWAWAA 796

Qy        823 CAACATCTTTGATGAAGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTT 882
              :|||    | |  ||  :|||| | :|| | |||  |  |     ||      |  |
Db        797 TWACAAAAATWAAAAAAAWTTTAAAAAWTTWTAAATTWAAAAAAWAATTTTAAAAAAAAT 856

Qy        883 TTATATTAAATCATAATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTC 942
              ||   ||  |  |:|||:|| | |  |:|| |  |     : | |||:| || | |
Db        857 TTTATTTTTAATTTWATWAATTTATAAWTAAAATTASAAAAAWTTTTTTWTTTATTTATT 916

Qy        943 TTTTAAAGAGGTAGGATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTAT 1002
              ||: |   :  |:   |     :||        :|||:|       :|:  ::|  |
Db        917 TTWAATTTWAATWTTTTTTWWAWAATAAAATAWAAAWAAAAAAATTWTWWAAWWTATAAA 976

Qy        1003 AATTCACCTCAAAGATACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAA 1062
               |       | ||: : | | |::    |::|: | ||:|:| |  :|   |  ||| |
Db         977 ATANWTAWTWAAWTWAAAAWTWWATWWAWWTWTAAAAWTWTATTTWWTWTAAAAAAAAT 1036

Qy        1063 TATTAGTAGATGTTGTTTTTGCTGCAAAAATAAAAATTTTT--CTGGTTGAAATAACTCA 1120
               ||||  |:    : |||::  |  |   |: |||||:|||   :  |   :|||| |
Db        1037 TATTTTTWTTAWAWTTTTWWTATATATTTAWTAAAATWTTTWAAWAWTATTWATAAATTT 1096

Qy        1121 AGGCCTCTAGCGTTTTCCTTTATCTTAAAATACAGGAAATAGCGATTGAAGTTAATTGAC 1180
               : :|    |:|  :| : |:  ::||    |||||   ||| :| |||::   |
Db        1097 TTATTWAWATAAATWTAWAWTTWTTWTTWWTATTAAAAATAAAAATTAWATTTAWWMWAT 1156

Qy        1181 ACTTAAGCAAATAGTCAACCTAACAGA 1207
               :   |   |: |  | ::  :|| | |
Db        1157 WTAAATWTAWTTTWTAWMWTWAAAAAA 1183


RESULT 14
BX437739/c
LOCUS       BX437739               1200 bp    mRNA    linear   EST 22-MAY-2003
DEFINITION  BX437739 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP008YG24
            3-PRIME, mRNA sequence.
ACCESSION   BX437739
VERSION     BX437739.1  GI:31018315
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1200)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 534.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0CAP008BD12NP1&cluster=534.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0CAP008BD12NP1.
FEATURES             Location/Qualifiers
     source          1. .1200
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0CAP008YG24"
                     /tissue_type="THYMUS"
                     /clone_lib="Homo sapiens THYMUS"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             2.1%;  Score 75.8;  DB 13;  Length 1200;
  Best Local Similarity   36.9%;  Pred. No. 4.2e-06;
  Matches  246;  Conservative   93;  Mismatches  322;  Indels    5;  Gaps    2;

Qy        542 TATTCAATTAATTCACTATTAAATAAACAGTTCTAAAAGGCTGTTTATTGGATGAATATT 601
              |:  ::::: ||||: |: || :::: ::  |:|:::   :  :|| ||  :    ::|
Db       1193 TMMMMHMHWTATTCMTTHMTATHHMMTHMTMTHTHWHTTHMMMWTTTTTTTMMMTTMMWT 1134

Qy        602 CGAAATTATCACATAATAATTGATGCTATTATTACTTGCTGTATTGGTATCAACTTTCAT 661
              : :::|| |   :::: : :| :|  |:||  ||  | |:|| |  :: : || |
Db       1133 MMMMTTTTTTTMHMMMTHTTMTMMTTMTMTTCWTTTTTMTHTTMTTTMMHTHMTTTTT 1074

Qy        662 GCTCTATACATGTAATATATTTCGAGTTAGACCTTAATTCAAGGTAATTTGTCTATTTAA 721
                :|:|:::   |: :  ||   : ||  :  ||: | :::  |: ||:  |     :
Db       1073 MTMMTMTMMMAYMAMMMTTTMTTMTTTTMMTTTTMTTMMMHTATHTTTHTACMTMMMMT 1014

Qy        722 TTATTATCTGAATAATATGTAATCGATTGCTTTGTGGTTATTTTTATGTTTGTTTCATTT 781
              | ||  | | |   : :: | :|   ||  |:       | |||:   : |  || ||:|
Db       1013 TMATMTTWTMATACMMMHMTTHTTTTTTTTTTHAMAMAMTTTTTHMTMAHMTMMTTAATWT 954

Qy        782 TTAATGACGGTGAGCTTGTGCATTCATATTTTTTATGATGACAACATCTTTGA----TGA 837
              ||  |              ||    ||| ||: : | |||     | |  || ||     | |
Db        953 TTTTTATANAATTTTTTTATAATTWATWAWATATATATATAAATAATATTATCAATTTTA 894

Qy        838 AGTATTTAAGATATTGTTAATGCATGAGGGGTTTGCGTGTATTTTTTATATTAAATCATA 897
              | | ||| | || || ||       | :    ||| : | ||:||      ||| |:| :|:
Db        893 AATTTTTTATATTTTTTTTTATTTTAWTWTTTTHATTTAWTTAAATAATTTAWTWWTW 834

Qy        898 ATAAAATCAACAATATATGTTATTTTGTGTCTTTTTATAGTGTTCTTTTAAAGAGGTAGG 957
              || ::||  | |:|||   ||| ||| | | || | |:|   |: |:|||||  :  ||
Db        833 ATTWWATATAAAWTATTAATTAATTTTTATTTTATAAWAAMWTWTTWTTAAAAWTATATC 774

Qy        958 ATGACCTAAAGGTCGCCTAAATATGGCGTAAATTGCCATTGCTATAATTCACCTCAAAGA 1017
              || |   :|   |    ||| |  | || |||  | |:||   ||   |||
Db        773 ATAAAAAWATAATATATAAAAMTTHAAATTATTTTTTATTTCAAWAAAA-ACAATAAANT 715
```

```
Qy       1018 TACACTATTGGCAAATTGACAAATATGTCACTTCGTATGAAACAATATTAGTAGATGTTG 1077
Db        714 AAAATAATWWTAAAMATAATAAYAAAATTAWTTTWWWAAAYAATWTATTATTATAAMAAA 655

Qy       1078 TTTTTGCTGCAAAAATAAAAATTTTTCTGGTTGAAATAACTCAAGGCCTCTAGCGTTTTC 1137
Db        654 MATAAAAATTAAAAAAWTAATWAAAAAMAAAAAAAATAATAAAAATAAWAAAAAAAAATA 595

Qy       1138 CTTTATCTTAAAATACAGGAAATAGCGATTGAAGTTAATTGACACTTAAGCAAATAGTCA 1197
Db         594 AAATAAAAAAAAAWTTATAAAHATAATATATAAATAWAAAMAAATTTNAAAMAATAATAA 535

Qy       1198 ACCTAA 1203
Db         534 AAAAAA 529
```

RESULT 15
BX366070
LOCUS BX366070 1201 bp mRNA linear EST 05-MAY-2003
DEFINITION BX366070 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DC007YA01 5-PRIME, mRNA sequence.
ACCESSION BX366070
VERSION BX366070.1 GI:30366975
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6169.r For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS5AA003ZH01QP1&cluster=6169.r. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS5AA003ZH01QP1.
FEATURES Location/Qualifiers
source 1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC007YA01"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

Query Match 2.1%; Score 75.8; DB 13; Length 1201;
Best Local Similarity 41.7%; Pred. No. 4.2e-06;
Matches 173; Conservative 50; Mismatches 189; Indels 3; Gaps 1;

```
Qy        556 ACTATTAAATAAACAGTTCTAAAAGGCTGTTTATTGGATGAATATTCGAAATTATCACAT 615
Db        778 AAWATAATATATAAWATATTATATATWTATWTWAWATATWAWTATATTWTATTTWTTTTT 837

Qy        616 AATAATTGATGCTATTATTACTTGCTGTATTGGTATCAACTTTCATGCTCTATACATGTA 675
Db        838 WTWTWTATAWGTATTDAAAAWTTAATATTTTTATAAWTWTAAWAAWATAWAATAAWTTWT 897

Qy        676 ATATATTTCGAGTTAGACCTTAATTCAAGGTAATTTGTCTATTTAATTATTATCTGAATA 735
Db        898 ATATAWATATATATATTTTTWWTTTTWAWAAWATTATWATATATATWTAAAATATWWAWT 957

Qy        736 ATATGTAATCGATTGCTTTGTGGTTATTTTTATGTTTGTTTCATTTTTAATGACGGTGAG 795
Db        958 AWATTATTTATAT---TWAATAAWTATATATWTTWDWTTAAAATWWTATATTWWATATWW 1014

Qy        796 CTTGTGCATTCATATTTTTTATGATGACAACATCTTTGATGAAGTATTTAAGATATTGTT 855
Db       1015 TATATAATATAATATATTTTATATATATAATATAATATWTATWATTWTTATATTATTTTT 1074

Qy        856 AATGCATGAGGGGTTTGCGTGTATTTTTTATATTAAATCATAATAAAATCAACAATATAT 915
Db       1075 ATWATATAAAWAAATAWWATATATAAWAAAATATATATAATATAWATWWAAWTTTTATAT 1134

Qy        916 GTTATTTTGTGTCTTTTTATAGTGTTCTTTTAAAGAGGTAGGATGACCTAAAGGT 970
Db       1135 TATWTWTTTAWTATATATATATTWTTTTATTATATAAAAAAAAAWAWATAAAKTT 1189
```

Search completed: July 2, 2004, 22:55:36
Job time : 8688 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 2, 2004, 22:56:15 ; Search time 77 Seconds
(without alignments)
1698.954 Million cell updates/sec

Title: US-10-780-347-2
Perfect score: 2343
Sequence: 1 MPLSKHQIEQLSKPLSDDSI.....DTQSKQDQKPQSSATSALSW 463

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	264.5	11.3	364	6	ADA35354	Ada35354 Acinetoba
2	211.5	9.0	376	6	ABM67333	Abm67333 Photorhab
3	152.5	6.5	1454	7	ADD93813	Add93813 Human apo
4	152.5	6.5	1503	2	AAW48845	Aaw48845 Human rec
5	152.5	6.5	1503	7	ADD93815	Add93815 Human HJ0
6	143	6.1	884	6	AAO19960	Aao19960 C elegans
7	138	5.9	2008	2	AAW22016	Aaw22016 Utrophin
8	136.5	5.8	7201	4	ABB71136	Abb71136 Drosophil
9	135	5.8	2013	4	AAB67964	Aab67964 Amino aci
10	134.5	5.7	1179	4	AAU37669	Aau37669 Streptoco
11	134.5	5.7	1179	4	AAM01107	Aam01107 CFE 110 p
12	134.5	5.7	1179	6	ABU46083	Abu46083 Protein e
13	131	5.6	1329	7	ADE15650	Ade15650 Human str
14	129.5	5.5	860	7	ADC38517	Adc38517 Human AML
15	129	5.5	1328	6	ABO14658	Abo14658 Novel hum
16	128	5.5	2017	4	ABG06301	Abg06301 Novel hum
17	127.5	5.4	624	6	ABM66947	Abm66947 Photorhab
18	127	5.4	1294	6	ABU33368	Abu33368 Protein e
19	127	5.4	1726	6	AAO16416	Aao16416 Human nuc
20	127	5.4	2278	6	ABR39818	Abr39818 Human SCA
21	126.5	5.4	752	5	AAU84329	Aau84329 Protein D
22	126.5	5.4	752	6	ABR92070	Abr92070 Human cer
23	126.5	5.4	785	6	ABR62257	Abr62257 GPBP-inte
24	126.5	5.4	893	6	ABR69651	Abr69651 Human CGD
25	126.5	5.4	1133	6	ABR62253	Abr62253 GPBP-inte
26	126.5	5.4	1133	6	ABR62252	Abr62252 GPBP-inte
27	126.5	5.4	1135	6	ABR62251	Abr62251 GPBP-inte
28	126	5.4	610	2	AAW68206	Aaw68206 M. catarr
29	124.5	5.3	1940	5	ABG79661	Abg79661 Invertebr
30	124.5	5.3	1940	6	ABU89712	Abu89712 Protein d
31	123	5.2	1935	7	ADD45207	Add45207 Rat Prote
32	122	5.2	1370	4	ABB68691	Abb68691 Drosophil
33	122	5.2	1971	4	ABG07508	Abg07508 Novel hum
34	121.5	5.2	1179	6	ABU01677	Abu01677 S. pneumo
35	121.5	5.2	3259	7	ADE56037	Ade56037 Human Pro
36	121.5	5.2	3259	7	ADE56033	Ade56033 Human Pro
37	121	5.2	1221	6	ABU20426	Abu20426 Protein e
38	120.5	5.1	1048	4	ABB59245	Abb59245 Drosophil
39	120.5	5.1	1294	4	ABB63502	Abb63502 Drosophil
40	120	5.1	1131	4	ABG07281	Abg07281 Novel hum
41	120	5.1	1497	4	ABG20153	Abg20153 Novel hum
42	120	5.1	1935	5	ABG31649	Abg31649 Amino aci
43	119.5	5.1	1066	4	AAG67418	Aag67418 Amino aci
44	119.5	5.1	1066	6	ABG72693	Abg72693 Fruitfly
45	119.5	5.1	1179	4	AAU38015	Aau38015 Streptoco

ALIGNMENTS

RESULT 1
ADA35354
ID ADA35354 standard; protein; 364 AA.
XX
AC ADA35354;
XX
DT 20-NOV-2003 (first entry)
XX
DE Acinetobacter baumannii protein #2515.
XX
KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
KW plant biocontrol agent.
XX
OS Acinetobacter baumannii.
XX
PN US6562958-B1.
XX
PD 13-MAY-2003.
XX
PF 04-JUN-1999; 99US-00328352.
XX
PR 09-JUN-1998; 98US-0088701P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton G, Bush D;
XX
DR WPI; 2003-576092/54.
DR N-PSDB; ADA31228.
XX
PT New Acinetobacter baumanii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX
PS Example; SEQ ID NO 6641; 328pp; English.
XX
CC The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents the amino acid sequence of an A.
CC baumannii protein.
XX
SQ Sequence 364 AA;

```
  Query Match             11.3%;  Score 264.5;  DB 6;  Length 364;
  Best Local Similarity   21.5%;  Pred. No. 2.3e-13;
  Matches   87;  Conservative   76;  Mismatches  152;  Indels   89;  Gaps   14;

Qy          8 IEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACLN 67
Db          5 ISELLKPINDSLLCG-----EDYSF---SNEFHEIKKA--RTQDDLLLDQGDWVAERKQA 54

Qy         68 KWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNP 127
Db         55 DWDFVAKSVSTLLIEKTKDIRLLTWVIEAWTHLN-GFEGMVKGITLTHTMLNQYWQDIHP 113

Qy        128 VLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVL--QLPLVGEVTFF-- 183
Db        114 II------EDDD-------LDQRIGLLQGLINQ--------LPMLLKKVPLTNTAPYYNL 152

Qy        184 ------------------DFQSAERKGEISQLKSMLTTTVAQERFAIQFKMENAKRCVTQ 225
Db        153 LDYDNFLYHENIRRKQTEEYESQSGPSELEQFDQAIFNT------SKTFQYSNYQEFNSV 206

Qy        226 LDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENAL-----VHLSGIKLAPKAEAKTVEQ 280
Db        207 LTEWNVLKQTLDHLMGLDSPSFAAIDSAFETIHSTLRKIYKAEAFGTGLAPSQEQAAV-- 264

Qy        281 EVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRDLAFHL 340
Db        265 ------ITTPSMENQVPVQIVSDQPMFQPQAQT---HL------------ANREQAMKV 302

Qy        341 LREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQN 384
Db        303 LQEIADYFQANEPHSPVSYMLQKTIKWSQMPLHEWLAQVIKDEH 346


RESULT 2
ABM67333
ID   ABM67333 standard; protein; 376 AA.
XX
AC   ABM67333;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Photorhabdus luminescens protein sequence #430.
XX
KW   Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW   detection; food; gene expression; plant; animal; microorganism; toxin;
KW   antibiotic; biopesticide; virulence factor; disease model; plague;
KW   whooping cough.
XX
OS   Photorhabdus luminescens.
XX
PN   WO200294867-A2.
XX
PD   28-NOV-2002.
XX
PF   07-FEB-2002; 2002WO-IB003040.
XX
PR   07-FEB-2001; 2001FR-00001659.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Duchaud E,  Taourit S,  Glaser P,  Frangeul L,  Kunst F,  Danchin A;
PI   Buchrieser C;
XX
DR   WPI; 2003-148459/14.
XX
PT   Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT   useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS   Claim 2; SEQ ID NO 430; 1205pp; French.
XX
CC   The invention relates to the isolation of genes and their encoded
CC   proteins from Photorhabdus luminescens. The isolated sequences are
CC   sources of probes and primers for detecting the genome of P. luminescens
CC   and related species; to study polymorphisms; for gene analysis and for
CC   detection/amplification of the genes. Antibodies (Ab) raised against the
CC   polypeptides encoded by the genes are used for detection/identification
CC   of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC   carry a gene-containing vector are used to select compounds that
CC   modulate, regulate, induce or inhibit expression of the genes in plants,
CC   animals or microorganisms other than P. luminescens and are able to alter
CC   response or sensitivity to toxins and antibiotics produced by P.
CC   luminescens. Cells transformed to express the genes are useful for
CC   recombinant production of the proteins, particularly toxins and
CC   antibacterials useful as insecticides, bactericides and fungicides. The
CC   genes, proteins, vectors containing the genes and Ab are also useful
CC   therapeutically (to treat microbial infection by bacteria or fungi that
CC   are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC   biopesticides. Other uses of the genes and the proteins are as virulence
CC   factors and for identifying targets of human diseases for which P.
CC   luminescens is a model (particularly plague and whooping cough). This
CC   sequence represents one of the isolated P. luminescens proteins
XX
SQ   Sequence 376 AA;

  Query Match             9.0%;  Score 211.5;  DB 6;  Length 376;
  Best Local Similarity   22.0%;  Pred. No. 6.5e-09;
  Matches   96;  Conservative   75;  Mismatches  159;  Indels  107;  Gaps   17;

Qy          3 LSKHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQ 62
Db         11 LAQTDIARFLQPIAADKPAGCDIEYE-----PIFEQINAARET-----------DDDFLQ 54

Qy         63 ------EACLNKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLAD 116
Db         55 DDTWGYETRQADWMKVSALCQEVLEKQSKDLQIACWLTQAQGEL-YGLAGISGGITLLAH 113

Qy        117 LSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFF----QLVGDSEESSILYAPVL 172
Db        114 LLETFWPVLYPPL--------DD--TEANSADARLGRLSWLDNQLVKQLDNLTLTDDGKL 163

Qy        173 QLPLVGEVTFFDFQSAERKGEISQLKSML--------------TTTVAQERFAIQFKME 217
Db        164 SLSVWQRVQYFE----QRAAVNSELRSALISDGYFGMAECDGSIRTTPAEQFNQLLTRAE 219

Qy        218 NAKRCVTQLDRLSA-LVSTKCHSLGSQSTNFGFAKSLLTRVENALVHLSGIKLAPKAEAK 276
Db        220 QVNKALAELQKIMVELLPDAGNSMSASTQRLQELVALIERFRD--------MVAPGS--- 268

Qy        277 TVEQEVAESS---VSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNMN 333
Db        269 --SQEYGNSNEAIAAGGDAPNGMNTTFL-----------TEDRAHHEMRAIAIGQMIN-- 313

Qy        334 RDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDALSTIFN 393
Db        314 ----------IANYFRQNEPTSPVPYLIERAARWANMGMAEWLEEMMSE----------N 353

Qy        394 AAGLNHLDQVLL-PEVS 409
Db        354 TSALQEIMRVMKGPEIN 370


RESULT 3
ADD93813
ID   ADD93813 standard; protein; 1454 AA.
XX
AC   ADD93813;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human apoptosis/nerve growth-related protein HJ06972.
XX
KW   Human; HJ06972; KIAA1079; AATYK homologue;
KW   apoptosis-associated tyrosine kinase; apoptosis;
KW   nerve cell differentiation; nerve cell proliferation; nerve cell growth;
```

```
KW   nerve cell division; drug screening; transgenic animal;
KW   apoptosis disorder; nerve regeneration; neurological disorder; hepatitis;
KW   hepatotropic; anti-inflammatory; virucide; cytostatic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .1443
FT                   /note= "This region is common to HJ06972 and its splice
FT                   variant AF06972"
FT   Region          1444. .1454
FT                   /label= HJ06972_specific_C-terminal_sequence
XX
PN   WO2003080836-A1.
XX
PD   02-OCT-2003.
XX
PF   26-MAR-2003; 2003WO-JP003713.
XX
PR   26-MAR-2002; 2002JP-00086843.
XX
PA   (KAZU-) KAZUSA DNA RES INST.
PA   (FUJI ) FUJISAWA PHARM CO LTD.
XX
PI   Miyoshi S,  Zenkoh J,  Satoh S,  Nishimura S;
XX
DR   WPI; 2003-865124/80.
DR   N-PSDB; ADD93812.
XX
PT   New polynucleotides encoding proteins useful for treating disorders
PT   associated with cell growth, division and death and hepatitis.
XX
PS   Claim 13; SEQ ID NO 2; 142pp; Japanese.
XX
CC   The invention relates to a splice variant of human HJ06972 (KIAA1079),
CC   designated AF06972 (ADD93815), and nucleic acids encoding it (ADD93814).
CC   Compared to the HJ06972 cDNA (ADD93812), AF06972 cDNA contains an
CC   additional 94 bp exon (ADD93816) between bases 4623-4624 of HJ06972,
CC   which results in a variant C-terminus (ADD93817) in AF06972. HJ06972 and
CC   its splice variant AF06972 are related to apoptosis-associated tyrosine
CC   kinase (AATYK), a protein which is involved in apoptosis and the
CC   differentiation or proliferation of nerve cells, and are likely to have
CC   similar activity. The invention also relates to vectors and host cells
CC   comprising AF06972 DNA sequences, the recombinant production of AF06972
CC   polypeptides, antibodies specific for AF06972, an immunoassay method
CC   using the antibodies, a method for detecting apoptosis regulatory
CC   activity, a method for evaluating a compound for its ability to induce
CC   apoptosis or nerve cell differentiation or proliferation, and a
CC   transgenic non-human animal model in which apoptosis, or nerve cell
CC   division and growth has been modified. AF06972 polypeptides and
CC   polynucleotides, and related compounds and methods of the invention may
CC   be used in the treatment of disorders associated with apoptosis, nerve
CC   cell growth and division, and hepatitis. The present sequence represents
CC   the human HJ06972 protein.
XX
SQ   Sequence 1454 AA;

  Query Match             6.5%;  Score 152.5;  DB 7;  Length 1454;
  Best Local Similarity   21.4%;  Pred. No. 0.0039;
  Matches  122;  Conservative   75;  Mismatches  189;  Indels  183;  Gaps   25;

Qy          3 LSKHQIEQLSKPLSD---DSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERD 59
              | | |:||   | ||   : :  :| ||   ||   : :   | ||  ||  |  | |
Db        370 LPKPQLEQ---PYSDRWYEVLQFCWLSPEK---RPAAEDVHRLLTYLRLQSQRDS--EVD 421

Qy         60 ALQEACLNKWKILSDSLYEQFSKTTRDIELISWFVAAQF------LLDTTLESAANSLEW 113
                |:     |  |  :   : |       ::  |   :       :|  |  |   | |:
Db        422 FEQQ-----WNALKPNTNSRDSSNNAAFPILDHFARDRLGREMEEVLTVTETSQGLSFEY 476

Qy        114 LADLSE-KHWD-----HLN----------PVLPVETLKSDDDKGKEREQADAKVKAFFQL 157
              : : ::  |:|     ||:          ||   |:  ||   ||:
Db        477 VWEAAKHDHFDERSRGHLDEGLSYTSIFYPVEVFESSLSDPGPGKQ------------- 522

Qy        158 VGDSEESSILYAPVLQLPLVGEVTFFD----------FQSAERKG---EISQLKSMLTT 203
                ||  :   |  |      | |  ||           |  |: |   |:    ::|||
Db        523 -DDSGQDVPLRVP-------GVVPVFDAHNLSVGSDYYIQLEEKSGSNLELDYPPALLTT 574

Qy        204 TVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVH 263
                        |:| :|   :| :|:|| | :   |   ||:  | :|     :::|
Db        575 ----------DMDNPERTGPELSQLTALRSVE---LEESSTDEDFFQSSTDPKDSSLPG 620

Qy        264 LSGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQ----TVSQHLH 319
                :   |::    :  :| :|    :||||  |  : : :   ||      |:|  |
Db        621 DLHVTSGPESPFNNIFNDVDKSE----DLPSHQ--KIFDLMELNGVQADFKPATLSSSLD 674

Qy        320 --------------------------AGNLSELGNLNNMNRDLAFHLLREVSDYFRQSE 352
                                       : ||    | : :|       ::|:|: |
Db        675 NPKESVITGHFEKEKPRKIFDSEPLCLSDNLMHQDNFDPLN-------VQELSENF---- 723

Qy        353 PHSPISFLLEKAIRWGYLSLPELLREMMSE-QNGDALSTIFNAAGLNHLD---------- 401
                || || :  | ||  | : :: :| :|      :   :  | ||
Db        724 -----LFLQEKNLLKGSLSSKEHINDLQTELKNAGFTEAMLETSCRNSLDTELQFAENKP 778

Qy        402 ----------------QVLL-----------PEVSTPTVGIESPQTPQAKPSVSDPRSV 433
                             |:|           |||  |    |: :||:  |  | |
Db        779 GMSLLQENVSTKGDDTDVMLTGDTLSTSLQSSPEVQVPPTSFETEETPRRVPPDSLPTQG 838

Qy        434 EEHVSQTSPVDTQSKQDQKPQSSATSALS 462
                |   :| : :|    :|  |  :  |::
Db        839 E---TQPTCLDVIVPEDCLHQDISPDAVT 864
```

```
RESULT 4
AAW48845
ID   AAW48845 standard; protein; 1503 AA.
XX
AC   AAW48845;
XX
DT   12-OCT-1998  (first entry)
XX
DE   Human receptor tyrosine kinase LMR2_h.
XX
KW   Receptor tyrosine kinase; LMR2_h; human; signal transduction; cancer;
KW   neurodegenerative disorder.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .23
FT                   /label= Sig_peptide
FT   Domain          43. .69
FT                   /label= TMD
FT                   /note= "transmembrane domain"
FT   Peptide         84. .98
FT                   /note= "immunogen"
FT   Peptide         163. .166
FT                   /note= "protein kinase conserved motif"
FT   Peptide         281. .283
FT                   /note= "protein kinase conserved motif"
FT   Peptide         329. .332
FT                   /note= "protein kinase conserved motif"
FT   Modified-site   409
FT                   /label= O-phosphorylated
FT   Modified-site   477
FT                   /label= O-phosphorylated
FT   Modified-site   552
FT                   /label= O-phosphorylated
FT   Peptide         684. .698
FT                   /note= "immunogen"
FT   Modified-site   1032
FT                   /label= O-phosphorylated
FT   Peptide         1098. .1113
FT                   /note= "immunogen"
FT   Modified-site   1100
FT                   /label= O-phosphorylated
FT   Modified-site   1105
```

```
FT                   /label= O-phosphorylated
FT   Modified-site   1365
FT                   /label= O-phosphorylated
FT   Peptide         1488. .1503
FT                   /note= "immunogen"
XX
PN   WO9822507-A2.
XX
PD   28-MAY-1998.
XX
PF   21-NOV-1997;   97WO-US022526.
XX
PR   22-NOV-1996;   96US-0031675P.
XX
PA   (SUGE-) SUGEN INC.
XX
PI   Joho KE,  Plowman GD;
XX
DR   WPI; 1998-312419/27.
DR   N-PSDB; AAV32452.
XX
PT   New isolated receptor tyrosine kinase genes - which are expressed in
PT   neuronal tissues and tumour cells, useful as targets for
PT   neurodegenerative disorders or cancers.
XX
PS   Example 1; Fig 7B; 87pp; English.
XX
CC   This is the amino acid sequence of human LMR2-h, deduced from a LMR2-h
CC   cDNA clone (see AAV32452). Novel rat, human and mouse LMR1, LMR2 and LMR3
CC   (see AAW48841-49) define a novel family of receptors that are
CC   structurally related to receptor tyrosine kinases (RTKs). They all share
CC   the distinct motifs that typically characterise this class of enzymes,
CC   but possess extremely short extracellular domains and have C-terminal
CC   tails of unprecented length among RTKs. Expression of LMR1 and LMR3 is
CC   highly restricted to neuronal tissues with minimal expression in other
CC   adult or embryonic organs or in human tumour cell lines. LMR2 expression
CC   is limited to adult neuronal tissues, but is also very abundantly
CC   expressed in other non-neuronal foetal tissues and in numerous tumour
CC   cell lines. Based on restricted expression of all 3 LMRs to adult
CC   neuronal tissues and the up regulation of LMR2 in a wide variety of
CC   tumour cell lines, these proteins may be critical targets for
CC   neurodegenerative disorders or cancer. Anti-LMR antibodies may be used
CC   for detecting neurodegenerative diseases or cancer
XX
SQ   Sequence 1503 AA;

  Query Match             6.5%;  Score 152.5;  DB 2;  Length 1503;
  Best Local Similarity   21.4%;  Pred. No. 0.0041;
  Matches  122;  Conservative   75;  Mismatches  189;  Indels  183;  Gaps   25;

Qy          3 LSKHQIEQLSKPLSD---DSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERD 59
              | | |:||   | ||   : :   :|  ||   ||   : :   | ||  ||  |  | | |
Db        370 LPKPQLEQ---PYSDRWYEVLQFCWLSPEK---RPAAEDVHRLLTYLRLQSQRDS--EVD 421

Qy         60 ALQEACLNKWKILSDSLYEQFSKTTRDIELISWFVAAQF------LLDTTLESAANSLEW 113
               |:     |  |  :   : |        ::  |   :       :|  |  |   | |:
Db        422 FEQQ-----WNALKPNTNSRDSSNNAAFPILDHFARDRLGREMEEVLTVTETSQGLSFEY 476

Qy        114 LADLSE-KHWD-----HLN----------PVLPVETLKSDDDKGKEREQADAKVKAFFQL 157
              : : ::  |:|     ||:          ||   |:  ||   ||:
Db        477 VWEAAKHDHFDERSRGHLDEGLSYTSIFYPVEVFESSLSDPGPGKQ-------------- 522

Qy        158 VGDSEESSILYAPVLQLPLVGEVTFFD----------FQSAERKG---EISQLKSMLTT 203
                || :   |  |        | |  ||            |  |: |   |:    ::|||
Db        523 -DDSGQDVPLRVP-------GVVPVFDAHNLSVGSDYYIQLEEKSGSNLELDYPPALLTT 574

Qy        204 TVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVH 263
                        |:| :|   :| :|:|| | :   |   ||:  | :|     :::|
Db        575 ----------DMDNPERTGPELSQLTALRSVE---LEESSTDEDFFQSSTDPKDSSLPG 620

Qy        264 LSGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQ----TVSQHLH 319
                :   |::   :  :| :|     :|||   |  : : :   ||     |:|  |
Db        621 DLHVTSGPESPFNNIFNDVDKSE----DLPSHQ--KIFDLMELNGVQADFKPATLSSSLD 674

Qy        320 --------------------------AGNLSELGNLNNMNRDLAFHLLREVSDYFRQSE 352
                                        : ||    | : :|       ::|:|: |
Db        675 NPKESVITGHFEKEKPRKIFDSEPLCLSDNLMHQDNFDPLN-------VQELSENF---- 723

Qy        353 PHSPISFLLEKAIRWGYLSLPELLREMMSE-QNGDALSTIFNAAGLNHLD---------- 401
                 || || :  | ||  | : :: :| :|      :   :  | ||
Db        724 -----LFLQEKNLLKGSLSSKEHINDLQTELKNAGFTEAMLETSCRNSLDTELQFAENKP 778

Qy        402 ---------------QVLL-----------PEVSTPTVGIESPQTPQAKPSVSDPRSV 433
                            |:|             |||  |    |: :||:  |  | |
Db        779 GLSLLQENVSTKGDDTDVMLTGDTLSTSLQSSPEVQVPPTSFETEETPRRVPPDSLPTQG 838

Qy        434 EEHVSQTSPVDTQSKQDQKPQSSATSALS 462
              |   :| : :|    :|   |  :  |::
Db        839 E---TQPTCLDVIVPEDCLHQDISPDAVT 864


RESULT 5
ADD93815
ID   ADD93815 standard; protein; 1503 AA.
XX
AC   ADD93815;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human HJ06972 splice variant AF06972.
XX
KW   Human; AF06972; HJ06972 splice variant; KIAA1079 splice variant;
KW   AATYK homologue; apoptosis-associated tyrosine kinase; apoptosis;
KW   nerve cell differentiation; nerve cell proliferation; nerve cell growth;
KW   nerve cell division; drug screening; transgenic animal;
KW   apoptosis disorder; nerve regeneration; neurological disorder; hepatitis;
KW   hepatotropic; anti-inflammatory; virucide; cytostatic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .1443
FT                   /note= "This region is common to HJ06972 and its splice
FT                   variant AF06972"
FT   Region          1444. .1503
FT                   /label= AF06972_specific_C-terminal_sequence
FT                   /note= "Corresponds to SEQ ID NO:6"
XX
PN   WO2003080836-A1.
XX
PD   02-OCT-2003.
XX
PF   26-MAR-2003; 2003WO-JP003713.
XX
PR   26-MAR-2002; 2002JP-00086843.
XX
PA   (KAZU-) KAZUSA DNA RES INST.
PA   (FUJI ) FUJISAWA PHARM CO LTD.
XX
PI   Miyoshi S,  Zenkoh J,  Satoh S,  Nishimura S;
XX
DR   WPI; 2003-865124/80.
DR   N-PSDB; ADD93814.
XX
PT   New polynucleotides encoding proteins useful for treating disorders
PT   associated with cell growth, division and death and hepatitis.
XX
PS   Claim 1; SEQ ID NO 4; 142pp; Japanese.
XX
CC   The invention relates to a splice variant of human HJ06972 (KIAA1079),
CC   designated AF06972 (ADD93815), and nucleic acids encoding it (ADD93814).
CC   Compared to the HJ06972 cDNA (ADD93812), AF06972 cDNA contains an
CC   additional 94 bp exon (ADD93816) between bases 4623-4624 of HJ06972,
CC   which results in a variant C-terminus (ADD93817) in AF06972. HJ06972 and
CC   its splice variant AF06972 are related to apoptosis-associated tyrosine
CC   kinase (AATYK), a protein which is involved in apoptosis and the
```

```
CC   differentiation or proliferation of nerve cells, and are likely to have
CC   similar activity. The invention also relates to vectors and host cells
CC   comprising AF06972 DNA sequences, the recombinant production of AF06972
CC   polypeptides, antibodies specific for AF06972, an immunoassay method
CC   using the antibodies, a method for detecting apoptosis regulatory
CC   activity, a method for evaluating a compound for its ability to induce
CC   apoptosis or nerve cell differentiation or proliferation, and a
CC   transgenic non-human animal model in which apoptosis, or nerve cell
CC   division and growth has been modified. AF06972 polypeptides and
CC   polynucleotides, and related compounds and methods of the invention may
CC   be used in the treatment of disorders associated with apoptosis, nerve
CC   cell growth and division, and hepatitis. The present sequence represents
CC   the human HJ06972 splice variant AF06972.
XX
SQ   Sequence 1503 AA;

  Query Match             6.5%;  Score 152.5;  DB 7;  Length 1503;
  Best Local Similarity   21.4%;  Pred. No. 0.0041;
  Matches  122;  Conservative   75;  Mismatches  189;  Indels  183;  Gaps   25;

Qy          3 LSKHQIEQLSKPLSD---DSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERD 59
              | | |:||   | ||    : :   :| ||   ||   : :   | ||  ||  |  | |
Db        370 LPKPQLEQ---PYSDRWYEVLQFCWLSPEK---RPAAEDVHRLLTYLRLQSQRDS--EVD 421

Qy         60 ALQEACLNKWKILSDSLYEQFSKTTRDIELISWFVAAQF------LLDTTLESAANSLEW 113
                |:     |  |  :   : |         ::  |   :        :|  |  |   | |:
Db        422 FEQQ-----WNALKPNTNSRDSSNNAAFPILDHFARDRLGREMEEVLTVTETSQGLSFEY 476

Qy        114 LADLSE-KHWD-----HLN----------PVLPVETLKSDDDKGKEREQADAKVKAFFQL 157
              : : ::  |:|      ||:          ||   |:  ||   ||:
Db        477 VWEAAKHDHFDERSRGHLDEGLSYTSIFYPVEVFESSLSDPGPGKQ-------------- 522

Qy        158 VGDSEESSILYAPVLQLPLVGEVTFFD----------FQSAERKG---EISQLKSMLTT 203
                || :   |  |       | |  ||             |  |: |   |:    ::|||
Db        523 -DDSGQDVPLRVP-------GVVPVFDAHNLSVGSDYYIQLEEKSGSNLELDYPPALLTT 574

Qy        204 TVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVH 263
                      |:| :|   :| :|:|| | :   |   ||:  | :|     :::|
Db        575 ----------DMDNPERTGPELSQLTALRSVE---LEESSTDEDFFQSSTDPKDSSLPG 620

Qy        264 LSGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQ----TVSQHLH 319
                :   |::    :  :| :|     :||||   |  : : :   ||      |:|  |
Db        621 DLHVTSGPESPFNNIFNDVDKSE----DLPSHQ--KIFDLMELNGVQADFKPATLSSSLD 674

Qy        320 --------------------------AGNLSELGNLNNMNRDLAFHLLREVSDYFRQSE 352
                                       : ||    | : :|       ::|:|: |
Db        675 NPKESVITGHFEKEKPRKIFDSEPLCLSDNLMHQDNFDPLN-------VQELSENF---- 723

Qy        353 PHSPISFLLEKAIRWGYLSLPELLREMMSE-QNGDALSTIFNAAGLNHLD---------- 401
                  || || :  | ||  | : :: :| :|      :   :  | ||
Db        724 -----LFLQEKNLLKGSLSSKEHINDLQTELKNAGFTEAMLETSCRNSLDTELQFAENKP 778

Qy        402 ----------------QVLL------------PEVSTPTVGIESPQTPQAKPSVSDPRSV 433
                             |:|             |||  |    |: :||:  |  | |
Db        779 GMSLLQENVSTKGDDTDVMLTGDTLSTSLQSSPEVQVPPTSFETEETPRRVPPDSLPTQG 838

Qy        434 EEHVSQTSPVDTQSKQDQKPQSSATSALS 462
              |   :| : :|    :|   |  :  |::
Db        839 E---TQPTCLDVIVPEDCLHQDISPDAVT 864


RESULT 6
AAO19960
ID   AAO19960 standard; protein; 884 AA.
XX
AC   AAO19960;
XX
DT   18-AUG-2003  (first entry)
XX
DE   C elegans LLW-4 protein.
XX
KW   Adult lifespan; glycolysis; mitochondrial respiratory chain; aging;
KW   gene therapy; cytochrome C1; NADH oxidoreductase; ATP synthase;
KW   cytochrome C oxidase; phosphoglucose isomerase; GTPase; LLW-1; LLW-2;
KW   LLW-3; LLW-4; HSF-1.
XX
OS   Caenorhabditis elegans.
XX
PN   WO2003000861-A2.
XX
PD   03-JAN-2003.
XX
PF   24-JUN-2002; 2002WO-US020247.
XX
PR   22-JUN-2001; 2001US-0300577P.
PR   25-JUN-2001; 2001US-0301052P.
PR   18-APR-2002; 2002US-0373975P.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Kenyon C,  Apfeld J,  Dillin A,  Garigan D,  Hsu AA;
PI   Lehrer-Graiwer J,  Murphy C;
XX
DR   WPI; 2003-229334/22.
XX
PT   Identifying a compound that modulates aging, comprises contacting the
PT   compound with a polypeptide comprising, e.g. NADH oxidoreductase or
PT   phosphoglucose isomerase, and determining the effect of the compound on
PT   the polypeptide.
XX
PS   Claim 1; Page 89; 112pp; English.
XX
CC   The present invention relates to a method of identifying a compound that
CC   modulates aging, involving contacting a compound with one of cytochrome
CC   C1, NADH oxidoreductase, ATP synthase, cytochrome C oxidase,
CC   phosphoglucose isomerase, GTPase, LLW-1, LLW-2, LLW-3, LLW-4 and HSF-1,
CC   and determining the functional effect upon the polypeptide. The method is
CC   useful identifying compounds that modulate aging and in regulating adult
CC   lifespan in eukaryotes, e.g. humans, animals and plants. The compound may
CC   also be used in diagnosis and therapy related to lifespan extension, life
CC   expectancy, and aging. The present sequence is the C. elegans LLW-4
CC   protein, which is one of the proteins involved in lifespan regulation and
CC   described in the invention
XX
SQ   Sequence 884 AA;

  Query Match             6.1%;  Score 143;  DB 6;  Length 884;
  Best Local Similarity   20.7%;  Pred. No. 0.012;
  Matches  103;  Conservative   99;  Mismatches  165;  Indels  130;  Gaps   25;

Qy         27 LEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACLNKW----KILSDSLYEQFSK 82
              |   |:   |     :| |:   |||          :|  | |:    :::: :|: |  :
Db        218 LRPDAYTEARITRRLALTSTATLSQR---------RERFLKKYSQIRQVVAQNLFGQQQR 268

Qy         83 TTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDKGK 142
               || :| :    :::  ::   |       |   | ||           | | |: ::  :
Db        269 LTRPVEDLVTETSSKLAVEEAPEETTQEETTTDASE----------VTTTKAVEEATE 317

Qy        143 E-REQADAKVKAFFQLVGDSEESSILYAPVLQ-------------LPLVGEVTFFD-FQ 186
              |  |:|    :|    |  ::|||  ::   ::              : |: |  |  : |:
Db        318 EVTEEATEATEA---PVATTKESSEMHVNTIRNMIRSASEKDLSKYVTLISEGKFSELFE 374

Qy        187 SAERK------------GEISQLKSMLTTTVAQ-------ERFAIQFKMENAKRCVTQL 226
               ||:|              ::::|| ::   :::       |:   :|:       |
Db        375 LAEQKKLTLTSKFDEKLSSKMAKLKDLINEALSEKEKSGEIEQAMEKFEKPEKSELVAME 434

Qy        227 DRLSALVSTKCHSLGSQSTNFGFAKSLLTR----VENALVHLSGIKLAPKAEAKTVEQE- 281
              |: :  | |   ||  :      | ::::|     |||   :    : |||| | |::|
Db        435 DKDTPAVFTISDSLKHKKAEAKLAHTIVSRNVVEAENA---IEKEVVEPKAEEKKVKEED 491

Qy        282 ---VAESSVSE---GELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNM--N 333
                 |||    |   |:||  |  : :|: |: ::              ||   |  :  :
Db        492 VKAVAEEKKEEKKPGKLP--MKIEKLEK-PVDTK-------------SENHELKKVLDD 534

Qy        334 RDLAFHLLREVSDYFRQSEPHSPISFLLEK---AIRWGYLSLPELLREMMSEQ---NGDA 387
              :: |  :  |: :   :::|    ||  |:   ||      :| | :|  :|:    |:
```

```
Db        535 KERALLVESEIKNTAEETKPKVE-SFKSEETTVAID----DMPALEKEESAEKKETTGEP 589

Qy        388 LSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPRSVEEHVSQTSPVDTQ- 446

Db        590 TTT----------------EAAVETTEASETPKPEAKPELLS--NLEDVLTLTTP-ETET 630

Qy        447 ---SKQDQKPQSSATSA 460

Db        631 IEGSGEREEPTTSAPAA 647


RESULT 7
AAW22016
ID   AAW22016 standard; protein; 2008 AA.
XX
AC   AAW22016;
XX
DT   03-OCT-1997  (first entry)
XX
DE   Utrophin truncated polypeptide.
XX
KW   Utrophin; minigene; Duchenne muscular dystrophy; gene therapy.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Region          238. .250
FT                   /note= "residues 238-250 (all encoded by codon NNN) may
FT                   comprise the sequence DKKSIIMYLTSL, absolutely conserved
FT                   in human, mouse and rat utrophins"
XX
PN   WO9722696-A1.
XX
PD   26-JUN-1997.
XX
PF   19-DEC-1996;   96WO-GB003156.
XX
PR   19-DEC-1995;   95GB-00025962.
PR   26-JUL-1996;   96GB-00015797.
PR   24-OCT-1996;   96GB-00022174.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Tinsley JM,  Davies KE;
XX
DR   WPI; 1997-341687/31.
DR   N-PSDB; AAT74665.
XX
PT   Nucleic acid encoding utrophin, truncated forms and related vectors -
PT   also transformed mammalian cells, used for alleviating symptoms of
PT   muscular dystrophy.
XX
PS   Claim 1; Fig 3; 78pp; English.
XX
CC   A truncated utrophin polypeptide (AAW22016) having utrophin function
CC   includes the actin-binding domain and dystrophin protein complex (DPC)
CC   binding domain of full-length utrophin (see also AAW22017), but lacks the
CC   rod domain. It is obtd. by expression of a utrophin minigene (AAT74665)
CC   comprising approx. the first 2 kb and the last 4 kb of the full-length
CC   utrophin coding sequence (see also AAT74666). Expression of the truncated
CC   utrophin significantly decreases the severity of the dystrophic muscle
CC   phenotype in an animal model, indicating usefulness in treatment of
CC   muscular dystrophy. The polypeptide can also be used to screen for
CC   substances that modulate utrophin binding to actin and/or the DPC
XX
SQ   Sequence 2008 AA;

  Query Match             5.9%;  Score 138;  DB 2;  Length 2008;
  Best Local Similarity   21.2%;  Pred. No. 0.1;
  Matches  105;  Conservative   72;  Mismatches  172;  Indels  146;  Gaps   19;

Qy          1 MPLSKHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDA 60

Db        420 MELQKKQLQQLSSWLA----------------------LTEERIQKMESPPLGDDLPS 455

Qy         61 LQEACLNKWKILSDSLYEQFSKTTRDIELI-----SWFVAAQFLLDTTLESAANSLEWLA 115

Db        456 LQK-LLQEHKSLQNDLEAEQVKVNSLTHMVVIVDENSGESATALLEDQLQKLGERWTAVC 514

Qy        116 DLSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLP 175

Db        515 RWTEERWNRL----------------------------------QEISILWQELLEEQ 538

Qy        176 LVGEVTFFDFQSAERKGEISQLKSMLTTTVAQERFAIQFKMENAKRCVTQLDRLSAL--- 232

Db        539 CLLEAWLTEKEEALNKVQTSNFKDQKELSVSVRRLAILKEDMEMKR--QTLDQLSEIGQD 596

Qy        233 ---------VSTKCHSLGSQSTNFGFAKSLLTRVENALVHLSGIKLAPKAEAKTVEQEVA 283

Db        597 VGQLLSNPKASKKMNSDSEELTQRW--DSLVQRLEDS--------------SNQVTQAVA 640

Qy        284 ESSVSEGELPSHMDTKHIERIPMASEQAQ------TVSQH-LHAGNLSELGNLNNMNRDL 336

Db        641 KLGMSQIPQKDLLETVHVREKGMVKKPKQELPPPLTKAEHAMQKRSTTELG--------- 691

Qy        337 AFHLLREVSDYFRQSEPH---------SPISFLLEKAIRWGYLSLPELLREMMSEQNGDA 387

Db        692 --ENLQELRDLTQEMEVHAEKLKWLNRTELEMLSDKS-----LSLPE--RDKISE----S 738

Qy        388 LSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPRSVEEHVSQTSPVDTQS 447

Db        739 LRTV------NMTWNKICREV--PTTLKECIQEPS---SVSQTR-IAAHPNVQKVVLVSS 786

Qy        448 KQDQKPQSSATSALS 462

Db        787 ASDIPVQSHRTSEIS 801


RESULT 8
ABB71136
ID   ABB71136 standard; protein; 7201 AA.
XX
AC   ABB71136;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 40200.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL15239.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 40200; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
```

```
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 7201 AA;

  Query Match             5.8%;  Score 136.5;  DB 4;  Length 7201;
  Best Local Similarity   22.0%;  Pred. No. 0.88;
  Matches  132;  Conservative   90;  Mismatches  178;  Indels  201;  Gaps   37;

Qy         30 SAFRPLRNE-FNVAQTA---LRKLS------QNPSAD---ERD---ALQEA---CLNKWK 70
Db       1263 SEFVDQKNDIFSMLQTAETELRSLTPLQTDPKNVSQDLKSKRDLNVQLQQASHQLLPKLH 1322

Qy         71 ILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDH------ 124
Db       1323 ALKSELAPLAAPDKRPI-LEKEVTEVEKMFFNTMEHVKDRVGYLEDYSAK-WNNYKTRLA 1380

Qy        125 -----LNPVLP--VETLKSDDDKGKEREQADAKVKAFFQLVGD-SEESSILYAPVLQL-P 175
Db       1381 ELQEWANKVAPKNIEALQSEDLTPEERV---VKVQAFKRILGDRMKQLDLLAADASELAP 1437

Qy        176 LVGEVTFFDFQSAERKGEISQLKSMLTT------------------------------- 203
Db       1438 KEGNIA----EAKRLKGEITKLQEVLSAINRNVDHQAQAVQEDLVNWQQFQAGLQQIKPA 1493

Qy        204 --------------------TVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQ 243
Db       1494 VEQSEVKVNNVVSKPISLEEAVAMQQNAQQFETQ----CQEQLDKLHG-ISNISHKMLCK 1548

Qy        244 STNFGFAKSLLTRVEN--ALVHLSGIKLAPKAEA-----KTVEQEVA--ESSVSEGE--- 291
Db       1549 -TN---APDELDAMHSRWTAVHENAKQASAKLEKLVANWKSFDADAAKLEDWVGQGEQQM 1604

Qy        292 --LPSHMDTKHIERI------------PMASEQAQTV---------SQHLHAGNLSEL-- 326
Db       1605 SRRPAVLNTPHIDKLEKELVKLKSFNNEISQQQAKLVTLGQNADQISLHLAPEGAAALKD 1664

Qy        327 ------GNLNNMNRDLAFHLLREVSDYF--RQSEPHSPISFLLEKAIRWGYLSLPELLRE 378
Db       1665 RVNQMKGKLQKLSEATRGH-INEVSDAIISRQDFNAKLVNF-----SNW-----MEQLRN 1713

Qy        379 MMS---EQNGDALST---IFNAAGLNHLDQ-----VLLPEVSTPTVGIESPQTPQAKPSV 427
Db       1714 QVTQVEEINPERVETSLHVIHALLQEHADKKPSFNAIYDEVKQLALG----ATPEESNAL 1769

Qy        428 SDP--------RSVEEHVSQTSPV------------DTQS------KQDQKPQSSATSAL 461
Db       1770 NDAYTALVVNYQNLETNMLQKKAALEKWTELLGWKNDTESHLNYLKHQLDKPEGPAAEEL 1829

Qy        462 S 462
Db       1830 S 1830


RESULT 9
AAB67964
ID   AAB67964 standard; protein; 2013 AA.
XX
AC   AAB67964;
XX
DT   29-JUN-2001  (first entry)
XX
DE   Amino acid sequence of utrophin B isoform minigene.
XX
KW   Utrophin; promoter; chromosome 6q24; exon 1B; dystrophin;
KW   muscle-specific transcription; muscular dystrophy.
XX
OS   Unidentified.
XX
PN   WO200125461-A1.
XX
PD   12-APR-2001.
XX
PF   04-OCT-2000; 2000WO-GB003800.
XX
PR   04-OCT-1999;   99GB-00023423.
XX
PA   (ISIS-) ISIS INNOVATION LTD.
XX
PI   Burton E,  Tinsley J,  Davies K;
XX
DR   WPI; 2001-273582/28.
DR   N-PSDB; AAF84673.
XX
PT   Novel nucleic acid comprising promoter for mouse, human utrophin genes,
PT   for promoting tissue-specific transcription of linked sequences and in
PT   screening for substances able to modulate utrophin promoter activity.
XX
PS   Disclosure; Fig 9; 76pp; English.
XX
CC   The present sequence is encoded by an utrophin B isoform minigene. The
CC   specification describes human and murine utrophin alternative promoters.
CC   Utrophin is a 395 kDa protein encoded by a gene located on chromosome
CC   6q24. The alternative promoter is highly regulated, and has little
CC   similarity to the synaptically expressed promoter. The alternative
CC   promoter drives transcription of a widely expressed unique fist exon that
CC   splice into a common full length mRNA at exon 3. This unique exon (called
CC   exon 1B) encodes a novel 31 peptide which may be involved in binding to
CC   the muscle membrane. Utrophin alternative promoters are useful for
CC   promoting tissue-specific, preferably muscle-specific transcription of an
CC   operably linked sequence of nucleotides. Utrophin exon 1B polynucleotides
CC   and polypeptides are useful in the manufacture of medicament for treating
CC   a dystrophin phenotype in a mammal. Up-regulation of utrophin expression
CC   may compensate for dystrophin loss in muscular dystrophy patients
XX
SQ   Sequence 2013 AA;

  Query Match             5.8%;  Score 135;  DB 4;  Length 2013;
  Best Local Similarity   21.2%;  Pred. No. 0.18;
  Matches  105;  Conservative   72;  Mismatches  172;  Indels  146;  Gaps   19;

Qy          1 MPLSKHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDA 60
Db        425 MELQKKQLQQLSSWLA----------------------LTEERIQKMESLPLGDDLPS 460

Qy         61 LQEACLNKWKILSDSLYEQFSKTTRDIELI-----SWFVAAQFLLDTTLESAANSLEWLA 115
Db        461 LQK-LLQEHKSLQNDLEAEQVKVNSLTHMVVIVDENSGESATALLEDQLQKLGERWTAVC 519

Qy        116 DLSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLP 175
Db        520 RWTEERWNRL----------------------------------QEISILWQELLEEQ 543

Qy        176 LVGEVTFFDFQSAERKGEISQLKSMLTTTVAQERFAIQFKMENAKRCVTQLDRLSAL--- 232
Db        544 CLLEAWLTEKEEALDKVQTSNFKDQKELSVSVRRLAILKEDMEMKR--QTLDQLSEIGQD 601

Qy        233 ---------VSTKCHSLGSQSTNFGFAKSLLTRVENALVHLSGIKLAPKAEAKTVEQEVA 283
Db        602 VGQLLSNPKASKKMNSDSEELTQRW--DSLVQRLEDS-------------SNQVTQAVA 645

Qy        284 ESSVSEGELPSHMDTKHIERIPMASEQAQ------TVSQH-LHAGNLSELGNLNNMNRDL 336
Db        646 KLGMSQIPQKDLLETVHVREQGMVKKPKQELPPPLTKAEHAMQKRSTTELG--------- 696

Qy        337 AFHLLREVSDYFRQSEPH---------SPISFLLEKAIRWGYLSLPELLREMMSEQNGDA 387
Db        697 --ENLQELRDLTQEMEVHAEKLKWLNRTELEMLSDKS-----LSLPE--RDKISE----S 743
```

```
Qy        388 LSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPRSVEEHVSQTSPVDTQS 447
Db        744 LRTV------NMTWNKICREV--PTTLKECIQEPS---SVSQTR-IAAHPNVQKVVLVSS 791

Qy        448 KQDQKPQSSATSALS 462
Db        792 ASDIPVQSHRTSEIS 806
```

RESULT 10
AAU37669
ID AAU37669 standard; protein; 1179 AA.
XX
AC AAU37669;
XX
DT 14-FEB-2002 (first entry)
XX
DE Streptococcus pneumoniae cellular proliferation protein #98.
XX
KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.
XX
OS Streptococcus pneumoniae.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS55528.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Example 3; SEQ ID NO 13262; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence represents an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1179 AA;

Query Match 5.7%; Score 134.5; DB 4; Length 1179;
Best Local Similarity 21.6%; Pred. No. 0.092;
Matches 100; Conservative 80; Mismatches 178; Indels 105; Gaps 24;

```
Qy          5 KHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEA 64
Db        674 KPELEQLQKEIAADE---ASLGSEEAALKTLQDQMAALTERLEAIK---SQGEQARIQEQ 727

Qy         65 CLNKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDH 124
Db        728 GLS-------LAYQQTSQQVEELETL-WKLQEE-EIDRLSEG-----DWQAD-KEKCQES 772

Qy        125 L--------NPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPL 176
Db        773 LATIASEKQNLEAEIEEIKSNKNAIQERYQ----------NLQEEVAQARLLKTEL 818

Qy        177 VGEVTFFDFQSAERKGEISQLKSMLTTTVAQERF--AIQFKMENAKRCVTQLDRLSALVS 234
Db        819 QGQKR-YEVADIERLG-----KELDNLNIEQEEIQRMLQEKVDNLEKVDTE------LLS 866

Qy        235 TKCHSLGSQSTNF--GFAKSL--LTRVENALVHLSG----------IKLAPKAEAKTV 278
Db        867 QQAEESKTQKTNLQQGLIRKQFELDDIEGQLDDIASHLDQARQQNEEWIRKQTRAEAK-- 924

Qy        279 EQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLS--ELGNLNNMNRDL 336
Db        925 KEKVSERL-------RHLQNQLTDQYQISYTEALEKAHELENLNLAEQEVQDLEKAIRSL 977

Qy        337 AFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDA---LSTIFN 393
Db        978 GPVNLEAIDQY---EEVHNRLDFL--NSQRDDILSAKNLLLETITEMNDEVKERFKSTFE 1032

Qy        394 AA----------GLNHLDQVLLPEVSTPTVGIESPQTPQAK 424
Db       1033 AIRESFKVTFKQMFGGGQAD-LILTEGDLLTAGVEISVQPPGK 1074
```

RESULT 11
AAM01107
ID AAM01107 standard; protein; 1179 AA.
XX
AC AAM01107;
XX
DT 02-OCT-2001 (first entry)
XX
DE CFE 110 protein sequence.
XX
KW Antibacterial; vaccine; gene therapy; bacterial cell wall viability; CFE;
KW CEG; Conserved Essential Gene; bacterial infection; antisense therapy;
KW antibiotic resistance.
XX
OS Streptococcus pneumoniae.
XX
PN WO200149721-A2.
XX
PD 12-JUL-2001.
XX
PF 29-DEC-2000; 2000WO-US035604.
XX
PR 30-DEC-1999; 99US-0174089P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Dougherty TJ, Pucci MJ, Dougherty BA, Davison DB, Bruccoleri RE;
PI Thanassi JA;
XX
DR WPI; 2001-496721/54.
DR N-PSDB; AAH90806.
XX
PT Nucleic acids encoding conserved essential genes involved in bacterial
PT replication which are potential targets for the treatment of antibiotic
PT resistant bacterial infections.
XX
PS Claim 27; Page 365-369; 380pp; English.

```
XX
CC   The present invention relates to nucleic acids (AAH90701-AAH90918)
CC   encoding polypeptides (AAM01002-AAM01114), which are essential for the
CC   viability of a bacterial cell wall. The acronym CFE stands for "CEG For
CC   Expression", where CEG stands for "Conserved Essential Gene". The nucleic
CC   acids are useful for detecting the presence of proteins essential for the
CC   viability of a bacterial cell wall in samples such as cells, tissues,
CC   biological fluids, blood, serum, nose, ear or throat swabs with ligands,
CC   and for detecting corresponding target nucleic acid molecules with
CC   complementary sequences. The nucleic acids are also useful for
CC   determining whether a genomic nucleotide sequence of interest is
CC   essential for viability of a bacterial cell or whether it resides within
CC   an operon, by integrating an exogeneous nucleotide sequence comprising a
CC   portion of an open reading frame of the genomic sequence of interest
CC   (comprising 200-500 base pairs) into the genomic sequence of interest
CC   which confers a selectable phenotype to the cell, and determining cell
CC   viability with a selection agent such as chloramphenicol. The nucleic
CC   acids and proteins are also useful as vaccines and for treating bacterial
CC   infections with gene therapy and antisense therapy. The nucleic acids
CC   also enable identification of targets suitable for the treatment of
CC   antibiotic resistant bacterial infections
XX
SQ   Sequence 1179 AA;

  Query Match             5.7%;  Score 134.5;  DB 4;  Length 1179;
  Best Local Similarity   21.4%;  Pred. No. 0.092;
  Matches   99;  Conservative   80;  Mismatches  180;  Indels  103;  Gaps   23;

Qy          5 KHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEA 64
              | ::||| | :: |       |  |::| : |:::       |  :    |  |:  :||
Db        674 KPELEQLQKEIAADE---ASLGSEEAALKTLQDQMAALTERLEAIK---SQGEQARIQEQ 727

Qy         65 CLNKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLAD-------L 117
               |:         |:| |:   ::| : | :  :  :|   |      :| ||       |
Db        728 GLS-------LAYQQTSQQVEELETL-WKLQEE-EIDRLSEG-----DWQADKEKCQESL 773

Qy        118 SEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLV 177
              :    |  |    :| :||: :  :|| |               : :  :  | :|:  |
Db        774 ATIASDKQNLEAEIEEIKSNKNAIQERYQ-------------NLQEEVAQARLLKTKLQ 819

Qy        178 GEVTFFDFQSAERKGEISQLKSMLTTTVAQERF--AIQFKMENAKRCVTQLDRLSALVST 235
              |:  ::    || |    | :   : ||     :| |::| ::   |:      |:|
Db        820 GQKR-YEVADIERLG-----KELDNLNIEQEEIQRMLQEKVDNLEKVDTE------LLSQ 867

Qy        236 KCHSLGSQSTNF--GFAKSL--LTRVENALVHLSG-----------IKLAPKAEAKTVE 279
              :     :| ||   |  :    |  :|  |  ::           |:   :|||| :
Db        868 QAEESKTQKTNLQQGLIRKQFELDDIEGQLDDIASHLDQARQQNEEWIRKQTRAEAK--K 925

Qy        280 QEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLS--ELGNLNNMNRDLA 337
              ::|:|          |:  :  ::  ::  :|   :  |   ||:  |: :|    | |
Db        926 EKVSERL-------RHLQNQLTDQYQISYTEALEKAHELENLNLAEQEVQDLEKAIRSLG 978

Qy        338 FHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDA---LSTIFNA 394
                |  :  |    | |: : ||   : |   ||   || | ::| | :      : | |
Db        979 PVNLEAIDQY---EEVHNRLDFL--NSQRDDILSAKNLLLETITEMNDEVKERFKSTFEA 1033

Qy        395 A-----------GLNHLDQVLLPEVSTPTVGIESPQTPQAK 424
                        |   | ::| |    | |:|   | |
Db       1034 IRESFKVTFKQMFGGGQAD-LILTEGDLLTAGVEISVQPPGK 1074


RESULT 12
ABU46083
ID   ABU46083 standard; protein; 1179 AA.
XX
AC   ABU46083;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #31610.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA49953.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 74007; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1179 AA;

  Query Match             5.7%;  Score 134.5;  DB 6;  Length 1179;
  Best Local Similarity   21.6%;  Pred. No. 0.092;
  Matches  100;  Conservative   80;  Mismatches  178;  Indels  105;  Gaps   24;

Qy          5 KHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEA 64
              | ::||| | :: |       |  |::| : |:::       |  :    |  |:  :||
Db        674 KPELEQLQKEIAADE---ASLGSEEAALKTLQDQMAALTERLEAIK---SQGEQARIQEQ 727

Qy         65 CLNKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDH 124
               |:         |:| |:   ::| : | :  :  :|   |      :| ||  ||  :
Db        728 GLS-------LAYQQTSQQVEELETL-WKLQEE-EIDRLSEG-----DWQAD-KEKCQES 772
```

```
Qy        125 L--------NPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPL 176
              |        |    :| :||: :  :|| |               : :  :  | :|:  |
Db        773 LATIASEKQNLEAEIEEIKSNKNAIQERYQ-------------NLQEEVAQARLLKTEL 818

Qy        177 VGEVTFFDFQSAERKGEISQLKSMLTTTVAQERF--AIQFKMENAKRCVTQLDRLSALVS 234
               |:   ::    || |     | :    : ||     :| |::| ::  |:      |:|
Db        819 QGQKR-YEVADIERLG-----KELDNLNIEQEEIQRMLQEKVDNLEKVDTE------LLS 866

Qy        235 TKCHSLGSQSTNF--GFAKSL--LTRVENALVHLSG-----------IKLAPKAEAKTV 278
              :    :| ||   |  :   |  :|  |  ::             |:   :||||
Db        867 QQAEESKTQKTNLQQGLIRKQFELDDIEGQLDDIASHLDQARQQNEEWIRKQTRAEAK-- 924

Qy        279 EQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLS--ELGNLNNMNRDL 336
              :::|:|          |:  :  ::  ::  :|   :  |   ||:  |: :|    | |
Db        925 KEKVSERL-------RHLQNQLTDQYQISYTEALEKAHELENLNLAEQEVQDLEKAIRSL 977

Qy        337 AFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDA---LSTIFN 393
                |  :  |    | |: : ||   :  |   ||   || .| ::| | :      : |
Db        978 GPVNLEAIDQY---EEVHNRLDFL--NSQRDDILSAKNLLLETITEMNDEVKERFKSTFE 1032

Qy        394 AA-----------GLNHLDQVLLPEVSTPTVGIESPQTPQAK 424
              |           |    | ::| |    | |:|    | |
Db       1033 AIRESFKVTFKQMFGGGQAD-LILTEGDLLTAGVEISVQPPGK 1074
```

RESULT 13
ADE15650
ID ADE15650 standard; protein; 1329 AA.
XX
AC ADE15650;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human structural and cytoskeleton-associated protein (SCAP) #29.
XX
KW human; structural and cytoskeleton-associated protein; SCAP;
KW arteriosclerosis; atherosclerosis; cirrhosis; hepatitis; myelofibrosis;
KW psoriasis; cancer; pneumonia; chronic bronchitis; yellow fever;
KW influenza; measles; mumps; HIV; human T lymphotropic virus; rabies;
KW gastroenteritis; encephalitis; rubella; epilepsy;
KW ischaemic cerebrovascular disease; stroke; cerebral neoplasm;
KW Alzheimer's disease; Pick's disease; Huntington's disease; dementia;
KW Parkinson's disease; amyotrophic lateral sclerosis; atrophy;
KW hereditary ataxia; multiple sclerosis; meningitis; brain abscess;
KW prion disease; Creutzfeldt-Jakob disease; insomnia; neurofibromatosis;
KW cerebral palsy; myasthenia gravis; anxiety.
XX
OS Homo sapiens.
XX
PN WO2003062391-A2.
XX
PD 31-JUL-2003.
XX
PF 16-JAN-2003; 2003WO-US001772.
XX
PR 18-JAN-2002; 2002US-0350702P.
PR 25-JAN-2002; 2002US-0351715P.
PR 15-FEB-2002; 2002US-0357402P.
PR 10-MAY-2002; 2002US-0379880P.
PR 17-MAY-2002; 2002US-0381599P.
PR 07-JUN-2002; 2002US-0387270P.
PR 19-JUL-2002; 2002US-0397125P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, Griffin JA, Richardson TW, Tang YT, Thangavelu K;
PI Forsythe IJ, Becha SD, Chawla NK, Hafalia AJA, Swarnakar A;
PI Marquis JP, Gorvad AE, Baughn MR, Lu DAM, Arvizu CS, Kable AE;
PI Lee SY, Ramkumar J, Jiang X, Jackson AA, Khare R, Elliott VS;
PI Bulloch SA, Xu Y, Lee S, Lehr-Mason PM;
XX
DR WPI; 2003-671468/63.
DR N-PSDB; ADE15681.
XX
PT New isolated polypeptides useful for treating e.g. cell proliferative
PT disorders, viral infections and neurological disorders.
XX
PS Claim 1; SEQ ID NO 29; 357pp; English.
XX
CC The invention comprises the amino acid and coding sequences of human
CC structural and cytoskeleton-associated proteins (SCAP). The SCAP DNA and
CC protein sequences of the invention are useful for the diagnosis and
CC treatment of: arteriosclerosis, atherosclerosis, cirrhosis, hepatitis,
CC myelofibrosis, psoriasis, primary cancer, pneumonia, chronic bronchitis,
CC yellow fever, influenza, measles, mumps, HIV, human T lymphotropic virus,
CC rabies, gastroenteritis, encephalitis, rubella, epilepsy, ischaemic
CC cerebrovascular disease, stroke, cerebral neoplasm, Alzheimer's disease,
CC Pick's disease, Huntington's disease, dementia, Parkinson's disease,
CC amyotrophic lateral sclerosis, atrophy, hereditary ataxias, multiple
CC sclerosis, meningitis, brain abscess, prion disease, Creutzfeldt-Jakob
CC disease, insomnia, neurofibromatosis, cerebral palsy, myasthenia gravis,
CC anxiety. The present amino acid sequence represents a human SCAP of the
CC invention.
XX
SQ Sequence 1329 AA;

Query Match 5.6%; Score 131; DB 7; Length 1329;
Best Local Similarity 22.0%; Pred. No. 0.22;
Matches 108; Conservative 69; Mismatches 182; Indels 132; Gaps 23;

```
Qy          3 LSKHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADE---RD 59
              |  | |      | : |    : :||    :   ||    :  |::   : |:    :|
Db        907 LKAHVQEVAQHNLKEASSASQFEELE-IVLKEKENELKRLEAMLKERESDLSSKTQLLQD 965

Qy         60 ALQEACLNKWKI--LSDSLYEQFSKTTRDIELIS-----------WFVAAQFLLDTTLE 105
                |  | | :|  |    |:| |       ||:              |       ||:  :
Db        966 VQDENKLFKSQIEQLKQQNYQQASSFPPHEELLKVISEREKEISGLWNE-----LDSLKD 1020

Qy        106 SAANSLEWLADLSEKHWDHLNPVLPVETLKSD--DDKGKEREQ--------ADAKVKAFF 155
              :  :  :   || ||:|: :  :   | :  |  :   |||:|        |   :|  |
Db       1021 AVEHQRKKNNDLREKNWEAMEALASTEKMLQDKVNKTSKERQQQVEAVELEAKEVLKKLF 1080

Qy        156 QLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSMLTTTVAQERFAIQFK 215
                |  |  |:: |  |                |:| :    : |  |: ::|   :: |
Db       1081 PKV--SVPSNLSYGEWL-------------HGFEKKAK----ECMAGTSGSEEVKVLEHK 1121

Qy        216 MENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVHLSGIKLAPKAEA 275
              :: |       | :  |:  :|             ||:|   |  |  |
Db       1122 LKEA-------DEMHTLLQLECEKY----------KSVLAETEGILQKLQ---------- 1154

Qy        276 KTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRD 335
              ::|||  :  |   |  ||   | ::    :|||              ||  | : |:|
Db       1155 RSVEQEENKWKVKVDE--SHKTIKQMQSSFTSSEQ-------------ELERLRSENKD 1198

Qy        336 LAFHLLREVSDYFRQSEPHSPISFLLEKA----------IRWGYLSLPELLREMMSEQN- 384
              :          :  |:   |  :  |||          :|      | | | :: :|||
Db       1199 I---------ENLRREREH--LEMELEKAEMERSTYVTEVRELKAQLNETLTKLRTEQNE 1247

Qy        385 -----GDALSTIFNAAGLNHLDQVLLPEVSTPTVGIE-SPQTPQAKPSVSDPRSVEEHVS 438
                  ||   :  |    | |  : : :  |  || |  :|: : |  :  ||   ::
Db       1248 RQKVAGD----LHKAQQSLELIQSKIVKAAGDTTVIENSDVSPETESSEKETMSVS--LN 1301

Qy        439 QTSPVDTQSKQ 449
              ||    || :|
Db       1302 QTV---TQLQQ 1309
```

RESULT 14
ADC38517
ID ADC38517 standard; protein; 860 AA.
XX
AC ADC38517;
XX
DT 18-DEC-2003 (first entry)
XX

DE Human AMLP1 myosin-tail motif consensus protein SEQ ID NO:866.
XX
KW human; angiomotin-like protein 1; AMLP1; cytostatic; gene therapy.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO2003037931-A2.
XX
PD 08-MAY-2003.
XX
PF 01-NOV-2002; 2002WO-US035129.
XX
PR 01-NOV-2001; 2001US-0334773P.
XX
PA (AMSH) AMERSHAM BIOSCIENCES SV CORP.
XX
PI Shannon M, Phan T;
XX
DR WPI; 2003-430501/40.
XX
PT New isolated nucleic acid molecule encoding a human angiomotin-like
PT protein, useful for treating or preventing a disorder associated with
PT decreased or increased expression or activity of AMLP1.
XX
PS Example 2; SEQ ID NO 866; 172pp; English.
XX
CC The present invention describes the human angiomotin-like protein 1
CC (AMLP1). human AMLP1 has cytostatic activity, and can be used in gene
CC therapy. The AMLP1 protein, nucleic acid molecules, antibodies, and
CC compositions of the present invention can be used for treating or
CC preventing a disorder associated with decreased or increased expression
CC or activity of AMLP1. The present sequence represents a consensus myosin-
CC tail motif amino acid sequence, which is used in an example from the
CC present invention.
XX
SQ Sequence 860 AA;

Query Match 5.5%; Score 129.5; DB 7; Length 860;
Best Local Similarity 18.5%; Pred. No. 0.15;
Matches 98; Conservative 87; Mismatches 208; Indels 137; Gaps 17;

```
Qy          7 QIEQLSKPLSD------------------DSICGVYLKLEKSAFRPLRNEFNVAQTALR 47
              ||||| |  :                    |||     |  || | : | :: :  |  |
Db        139 QIEQLQKQKAKAEKEKSQLQAEVDDLLAQLDSITKAKLNAEKKA-KQLESQLSELQVKLD 197

Qy         48 KLSQNPSADERDALQEACLNKWKILSDSLYEQFSKTTRDIE----LISWFVAAQFLLDTT 103
              :| :                ||         |  : |  || :|     :|     :  |::
Db        198 ELQRQ------------LNDLTSQKSRLQSENSDLTRQLEEAEAQVSNLSKLKSQLESQ 244

Qy        104 LESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEE 163
              || |  |||       |:  :  |    :  |: | |  :|: : ::: ||  :     :
Db        245 LEEAKRSLE------EESRERANLQAQLRQLEHDLDSLREQLEEESEAKAELE-----RQ 293

Qy        164 SSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSMLTTTVAQ---------------- 207
              |   | : |        : |: : | |  |: :||  |   :::
Db        294 LSKANAEIQQWR-----SKFESEGALRAEELEELKKKLNQKISELEEAAEAANAKCDSLE 348

Qy        208 --------ERFAIQFKMENAKRCVTQL-----------------DRLSALVSTKCHSLG 241
                      |   :| ::| |    ::|                 | | | : |
Db        349 KTKSRLQSELEDLQIELERANAAASELEKKQKNFDKILAEWKRKVDELQAELDTAQREAR 408

Qy        242 SQSTNFGFAKSLLTRVENALVHLSGIKLAPKAEAKTVEQEVAE--SSVSEGELPSHMDTK 299
              : ||        | |::| |  |      | : | | :: |: :    : ||    |   |
Db        409 NLSTE-------LFRLKNELEELKDQVEALRRENKNLQDEIHDLTDQLGEGGRNVHELEK 461

Qy        300 HIERIPMASEQAQTVSQHLHAG-NLSELG------NLNNMNRDLAFHLLREVSDYFRQSE 352
                |:   :: |   :   |  | |         |: :  ::    |  :  ::    :
Db        462 ARRRLEAEKDELQAALEEAEAALELEESKVLRAQVELSQIRSEIERRLAEKEEEFENTRK 521

Qy        353 PHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDALSTIFNAAGLNHLDQVLLPEVSTPT 412
              |       ::||         |   : :|  | | :         : | : |  :::
Db        522 NH-------QRAIE-------SLQATLEAETKGKAEA--------SRLKKKLEGDINELE 559
```

Qy 413 VGIESPQTPQAKPSVSDPRSVEEHVSQTSPVDTQSKQDQKPQSSATSALS 462
: :: | : ::|::: | : || :::|: : | |:
Db 560 IALDHAN----KANAEAQKNVKKYQQQVKELQTQVEEEQRAREDAREQLA 605

RESULT 15
ABO14658
ID ABO14658 standard; protein; 1328 AA.
XX
AC ABO14658;
XX
DT 25-AUG-2003 (first entry)
XX
DE Novel human protein #31.
XX
KW Human; NOV; gene therapy; endocrine related disease; diabetes;
KW metabolism-related disease; obesity; central nervous system disorder;
KW Alzheimer's disease; Parkinson's disease; epilepsy; multiple sclerosis;
KW schizophrenia; depression; autoimmune disorder; inflammatory disorder;
KW psoriasis; allergy; lupus erythematosus; asthma; cancer;
KW inflammatory bowel disease; rheumatoid arthritis; osteoarthritis;
KW colon cancer; lung cancer; liver cancer; breast cancer; ovarian cancer;
KW prostate cancer; brain cancer; melanoma; liver disease; liver cirrhosis;
KW lung disease; emphysema; obstructive pulmonary disease; haemophilia;
KW stroke; infection.
XX
OS Homo sapiens.
XX
PN WO2003023002-A2.
XX
PD 20-MAR-2003.
XX
PF 09-SEP-2002; 2002WO-US028539.
XX
PR 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318130P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 17-APR-2002; 2002US-0373212P.
PR 06-SEP-2002; 2002US-00236177.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Spytek KA, Patturajan M, Gorman L, Li L, Anderson DW, Zhong M;
PI Gerlach VL, Vernet CAM, Ellerman K, Berghs C, Rothenberg ME, Guo X;
PI Shimkets RA, Leach MD, Catterton E, Kekuda R, Ji W, Miller CE;
PI Rieger DK, Taupier RJ, Shenoy SG, Liu X, Padigaru M, Alsobrook JP;
PI Lepley DM, Edinger SR, Burgess CE;
XX
DR WPI; 2003-313242/30.
DR N-PSDB; ACD19351.
XX
PT New cytoplasmic, nuclear membrane bound or secreted polypeptides (NOVX)
PT and polynucleotides, useful in gene therapy, e.g. for treating or
PT preventing obesity, multiple sclerosis, allergy, cancers, hemophilia,
PT stroke or infections.
XX
PS Claim 1; Page 151; 586pp; English.
XX
CC The invention describes a new isolated polypeptide (NOVX). The NOVX
CC polypeptide, nucleic acid and antibody are useful as therapeutics,
CC particularly in the manufacture of a medicament for treating a syndrome
CC associated with a human disease, which includes a pathology associated

CC with NOVX polypeptide. The DNA encoding the protein is useful in gene
CC therapy for treating the disease or condition. In particular, the NOVX
CC polypeptide or polynucleotide is useful for treating endocrine/
CC metabolism-related diseases (e.g. obesity or diabetes), central nervous
CC system disorders (e.g. Alzheimer's disease, Parkinson's disease,
CC epilepsy, multiple sclerosis, schizophrenia or depression), autoimmune
CC and inflammatory disorders (e.g. psoriasis, allergy, lupus erythematosus,
CC asthma, inflammatory bowel disease, rheumatoid arthritis or
CC osteoarthritis), cancers (e.g. colon, lung, liver, breast, ovarian,
CC prostate or brain cancers, or melanoma), liver diseases (e.g. liver
CC cirrhosis), lung diseases (emphysema or obstructive pulmonary disease),
CC haemophilia, stroke, or infections (e.g. viral, bacterial or parasitic).
CC These are also useful in developing powerful assay system for functional
CC analysis of various human disorders, as well as in diagnostic
CC applications, and for monitoring the effects of drugs during clinical
CC trials. This is the amino acid sequence of a novel human NOV protein
XX
SQ Sequence 1328 AA;

Query Match 5.5%; Score 129; DB 6; Length 1328;
Best Local Similarity 22.0%; Pred. No. 0.32;
Matches 108; Conservative 69; Mismatches 182; Indels 132; Gaps 23;

```
Qy          3 LSKHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADE---RD 59
              |  |  |    | : |   : :||    :   ||    :  |::   : |:    :|
Db        906 LKAHVQEVAQHNLKEASSASQFEELE-IVLKEKGNELKRLEAMLKERESDLSSKTQLLQD 964

Qy         60 ALQEACLNKWKI--LSDSLYEQFSKTTRDIELIS-----------WFVAAQFLLDTTLE 105
                |  | | :|  |    |:| |      ||:            |       ||:  :
Db        965 VQDENKLFKSQIEQLKQQNYQQASSFPPHEELLKVISEREKEISGLWNE-----LDSLKD 1019

Qy        106 SAANSLEWLADLSEKHWDHLNPVLPVETLKSD--DDKGKEREQ--------ADAKVKAFF 155
              :  :  :   || ||:|: :  :   | :  |  :   |||:|        |   :|  |
Db       1020 AVEHQRKKNNDLREKNWEAMEALASTEKMLQDKVNKTSKERQQQVEAVELEAKEVLKKLF 1079

Qy        156 QLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSMLTTTVAQERFAIQFK 215
                |  |  |:: |   |                |:| :    : |  |: ::|   :: |
Db       1080 PKV--SVPSNLSYGEWL------------HGFEKKAK----ECMAGTSGSEEVKVLEHK 1120

Qy        216 MENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVHLSGIKLAPKAEA 275
              :: |       | :  |:  :|             ||:|   |  |  |
Db       1121 LKEA-------DEMHTLLQLECEKY----------KSVLAETEGILQKLQ---------- 1153

Qy        276 KTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRD 335
              ::|||  :  |   |  ||   | ::   :|||              ||   | : |:|
Db       1154 RSVEQEENKWKVKVDE--SHKTIKQMQSSFTSSEQ-------------ELERLRSENKD 1197

Qy        336 LAFHLLREVSDYFRQSEPHSPISFLLEKA----------IRWGYLSLPELLREMMSEQN- 384
              :        :  |:   |  :   ||||          :|     | | | :: :|||
Db       1198 I---------ENLRREREH--LEMELEKAEMERSTYVTEVRELKAQLNETLTKLRTEQNE 1246

Qy        385 -----GDALSTIFNAAGLNHLDQVLLPEVSTPTVGIE-SPQTPQAKPSVSDPRSVEEHVS 438
                  ||    :  |     | |  : : :  |  || |  :|: : |  :  ||   ::
Db       1247 RQKVAGD----LHKAQQSLELIQSKIVKAAGDTTVIENSDVSPETESSEKETMSVS--LN 1300

Qy        439 QTSPVDTQSKQ 449
              ||    || :|
Db       1301 QTV---TQLQQ 1308
```

Search completed: July 3, 2004, 02:44:29
Job time : 81 secs

OM protein - protein search, using sw model

Run on: July 3, 2004, 02:42:04 ; Search time 24 Seconds
(without alignments)
995.951 Million cell updates/sec

Title: US-10-780-347-2
Perfect score: 2343
Sequence: 1 MPLSKHQIEQLSKPLSDDSI.....DTQSKQDQKPQSSATSALSW 463

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	264.5	11.3	364	4	US-09-328-352-6641	Sequence 6641, Ap
2	197	8.4	681	4	US-09-252-991A-25690	Sequence 25690, A
3	152.5	6.5	1503	3	US-08-976-255-14	Sequence 14, Appl
4	138	5.9	2008	4	US-09-091-501B-8	Sequence 8, Appli
5	137.5	5.9	357	4	US-09-252-991A-17293	Sequence 17293, A
6	126	5.4	610	4	US-09-336-447A-11	Sequence 11, Appl
7	123	5.2	949	4	US-09-252-991A-30776	Sequence 30776, A
8	120.5	5.1	999	2	US-08-770-301A-1	Sequence 1, Appli
9	120.5	5.1	999	3	US-09-175-581-1	Sequence 1, Appli
10	119.5	5.1	1066	3	US-09-541-782-8	Sequence 8, Appli
11	119.5	5.1	1066	4	US-09-723-820-8	Sequence 8, Appli
12	119.5	5.1	1066	4	US-10-270-085-8	Sequence 8, Appli
13	118.5	5.1	806	4	US-09-976-594-321	Sequence 321, App
14	115.5	4.9	2468	4	US-09-976-594-726	Sequence 726, App
15	114.5	4.9	431	4	US-09-134-000C-5352	Sequence 5352, Ap
16	114.5	4.9	955	1	US-08-006-676B-1	Sequence 1, Appli
17	114.5	4.9	955	1	US-08-282-845-2	Sequence 2, Appli
18	114.5	4.9	955	2	US-08-428-414A-3	Sequence 3, Appli
19	114.5	4.9	955	5	PCT-US94-00324-1	Sequence 1, Appli
20	114	4.9	1939	4	US-09-310-187A-1	Sequence 1, Appli
21	112.5	4.8	1130	4	US-09-976-594-280	Sequence 280, App
22	110.5	4.7	381	4	US-09-216-393B-8	Sequence 8, Appli
23	110.5	4.7	675	1	US-08-317-522A-9	Sequence 9, Appli
24	110.5	4.7	675	1	US-08-439-818A-9	Sequence 9, Appli
25	110.5	4.7	675	2	US-08-751-965-9	Sequence 9, Appli
26	110.5	4.7	675	2	US-08-738-975-9	Sequence 9, Appli
27	110.5	4.7	675	2	US-08-728-626-9	Sequence 9, Appli
28	110.5	4.7	675	3	US-08-808-599A-9	Sequence 9, Appli
29	110.5	4.7	1262	4	US-09-198-452A-97	Sequence 97, Appl
30	110	4.7	396	4	US-09-252-991A-23803	Sequence 23803, A
31	109.5	4.7	959	4	US-09-543-681A-6879	Sequence 6879, Ap
32	108.5	4.6	2125	4	US-09-919-172-29	Sequence 29, Appl
33	108	4.6	298	4	US-09-252-991A-26358	Sequence 26358, A
34	108	4.6	1087	4	US-09-914-259-12	Sequence 12, Appl
35	107.5	4.6	1231	4	US-08-714-741-41	Sequence 41, Appl
36	106.5	4.5	715	2	US-08-849-212-6	Sequence 6, Appli
37	106.5	4.5	1234	4	US-09-592-054-8	Sequence 8, Appli
38	106.5	4.5	2409	6	5180808-2	Patent No. 5180808
39	106	4.5	830	4	US-09-064-199-19	Sequence 19, Appl
40	106	4.5	1106	4	US-09-064-199-17	Sequence 17, Appl
41	106	4.5	1130	2	US-08-519-547A-6	Sequence 6, Appli
42	106	4.5	1130	4	US-09-064-199-18	Sequence 18, Appl
43	106	4.5	1207	4	US-09-064-199-16	Sequence 16, Appl
44	106	4.5	1481	2	US-08-616-844-40	Sequence 40, Appl
45	106	4.5	1481	2	US-08-599-654-40	Sequence 40, Appl

ALIGNMENTS

RESULT 1
US-09-328-352-6641
; Sequence 6641, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6641
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6641

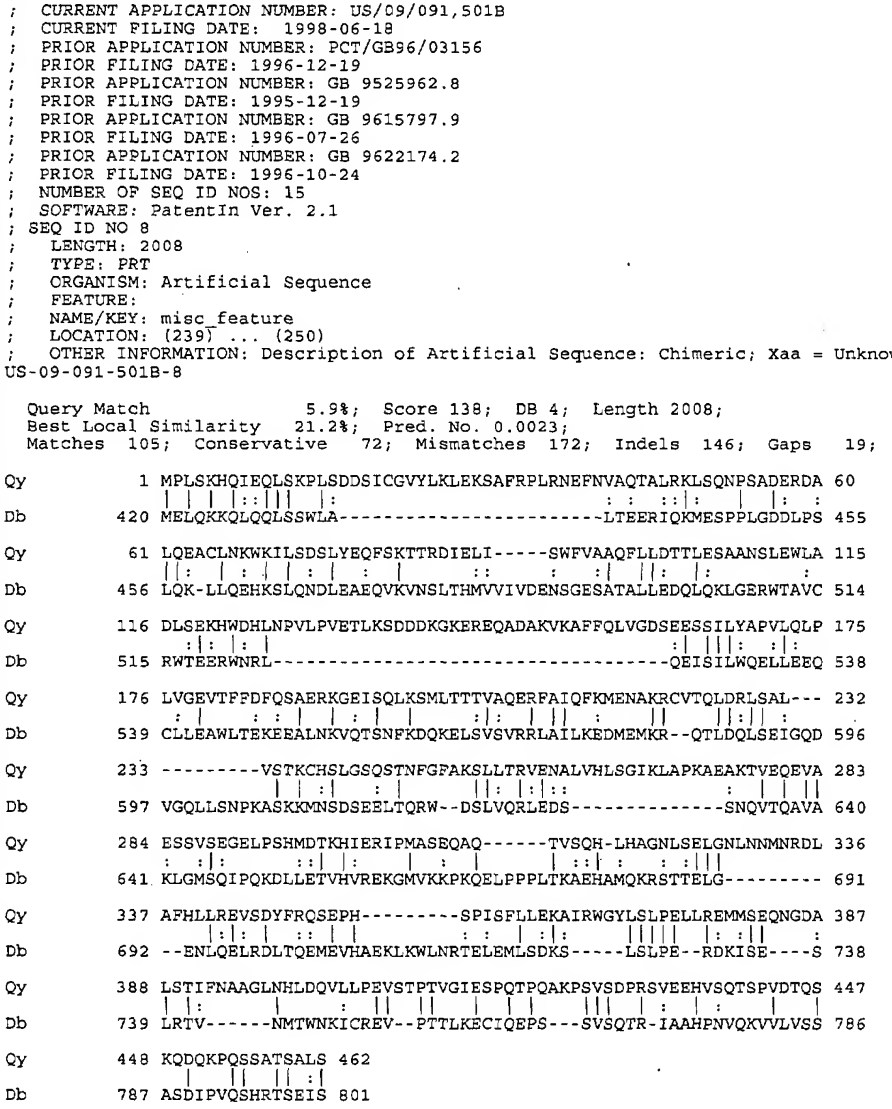
Query Match 11.3%; Score 264.5; DB 4; Length 364;
Best Local Similarity 21.5%; Pred. No. 1.7e-16;
Matches 87; Conservative 76; Mismatches 152; Indels 89; Gaps 14;

```
Qy          8 IEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACLN 67
              | :| ||::|  :||     |  :|    |||:  : |  :   :   |: | :|
Db          5 ISELLKPINDSLLCG-----EDYSF---SNEFHEIKKA--RTQDDLLLDQGDWVAERKQA 54

Qy         68 KWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNP 127
               |  :: |:    : |:|| |::| : |   |:   |    :    : ::|  ::|
Db         55 DWDFVAKSVSTLLIEKTKDIRLLTWVIEAWTHLN-GFEGMVKGITLTHTMLNQYWQDIHP 113

Qy        128 VLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVL--QLPLVGEVTFF-- 183
              ::       |||         | ::     |:          |:|  ::||     ::
Db        114 II------EDDD-------LDQRIGLLQGLINQ--------LPMLLKKVPLTNTAPYYNL 152

Qy        184 -----------------DFQSAERKGEISQLKSMLTTTVAQERFAIQFKMENAKRCVTQ 225
                              :::|     |: |    :  |      :  |:  | :   :
Db        153 LDYDNFLYHENIRRKQTEEYESQSGPSELEQFDQAIFNT------SKTFQYSNYQEFNSV 206

Qy        226 LDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENAL-----VHLSGIKLAPKAEAKTVEQ 280
              |   : |  |  | :|  | :|    |    : : |          |  |||  |   |
Db        207 LTEWNVLKQTLDHLMGLDSPSFAAIDSAFETIHSTLRKIYKAEAFGTGLAPSQEQAAV-- 264

Qy        281 EVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRDLAFHL 340
                ::   : : :  : :   ||   ||||   ||              ||: |  :
Db        265 ------ITTPSMENQVPVQIVSDQPMFQPQAQT---HL-------------ANREQAMKV 302

Qy        341 LREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQN 384
              |:|::|||: :|||||:|::|:| |:|  : | | | ::: :::
Db        303 LQEIADYFQANEPHSPVSYMLQKTIKWSQMPLHEWLAQVIKDEH 346
```

```
RESULT 2
US-09-252-991A-25690
; Sequence 25690, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 25690
;   LENGTH: 681
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25690

  Query Match             8.4%;  Score 197;  DB 4;  Length 681;
  Best Local Similarity   20.9%;  Pred. No. 1.1e-09;
  Matches   87;  Conservative   79;  Mismatches  156;  Indels   94;  Gaps   16;

Qy          1 MPLSKH--QIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADER 58
              :||| :   :| | :|:     ||  |: :    |            ||:| :   :
Db        318 LPLSGNALSLEVLLEPIDPGQPCGPSLRYDPDYDR-----------LRELRREDDSSLP 365

Qy         59 DALQEACLNK--WKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLAD 116
                : :|   :  | :     :  : ::|: | :|   | :|    |     :|  ||:
Db        366 TGVWQAEAKRADWAAVEQLASDLLQRRSKDLMLAAWLGEA-WLQRGGLGGLQRALVLLAE 424

Qy        117 LSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPL 176
              | |:: : ::|         | |:       |  :: : :|:             :|||
Db        425 LCERYPEEVHP------QAQDGDQSWRVPPIDWLLRRYAELLH-----------TRLPL 466

Qy        177 VG-----EVTFFDFQSAERK----GEISQLKSMLTTTVAQERFAIQFKMENAKR------ 221
              :|     |:| : :|  :|:    |:     |: |      |::     |:: | |
Db        467 MGQGAFAEITLYAWQRLQRQQVASGDSKSAKAALEAAQLQQK-----KLDEALRAEPLVQ 521

Qy        222 ----------CVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVHLSGIKLAP 271
                       |  || ||      :|  ||  : :    : :: :    |     | : |
Db        522 WQRKQASLLACQQQLQRLEQWCD-RC--LGELAPSCQPLREVIAQWLALLKEF--IAMHP 576

Qy        272 KAEAKTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNN 331
              :|     :  |||:  :||     |    | :|  ::      :
Db        577 QAPLPEEQPPVAEADAAEG------DADGEESVPASAPSGPAGAP--------------- 615

Qy        332 MNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDA 387
               :|: |:  |  ::||  ::|||||: :|:::|: ||   | ||| |::   | |:
Db        616 TSREDAYRQLLLIADYLARTEPHSPVPYLIKRAVEWGNKPLSELLAELI---NADS 668


RESULT 3
US-08-976-255-14
; Sequence 14, Application US/08976255
; Patent No. 6136581
;  GENERAL INFORMATION:
;    APPLICANT:  Jono, Keith E.
;    APPLICANT:  Plowman, Gregory
;    TITLE OF INVENTION:  KINASE GENES AND USES
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ for Windows 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/976,255
;      FILING DATE:  No. 6136581ember 21, 1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/031,675
;      FILING DATE:  No. 6136581ember 22, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  229/182
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1503 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Protein
US-08-976-255-14

  Query Match             6.5%;  Score 152.5;  DB 3;  Length 1503;
  Best Local Similarity   21.4%;  Pred. No. 6.1e-05;
  Matches  122;  Conservative   75;  Mismatches  189;  Indels  183;  Gaps   25;

Qy          3 LSKHQIEQLSKPLSD---DSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERD 59
              | | |:||   | ||   : :   :| ||   ||   : :   | ||  ||  |  | |
Db        370 LPKPQLEQ---PYSDRWYEVLQFCWLSPEK---RPAAEDVHRLLTYLRLQSQRDS--EVD 421

Qy         60 ALQEACLNKWKILSDSLYEQFSKTTRDIELISWFVAAQF------LLDTTLESAANSLEW 113
                |:     |  |  :   : :|        ::  |   :       :|  |  |   | |:
Db        422 FEQQ-----WNALKPNTNSRDSSNNAAFPILDHFARDRLGREMEEVLTVTETSQGLSFEY 476

Qy        114 LADLSE-KHWD-----HLN----------PVLPVETLKSDDDKGKEREQADAKVKAFFQL 157
              : : ::  |:|     ||:          ||   |:  ||   ||:
Db        477 VWEAAKHDHFDERSRGHLDEGLSYTSIFYPVEVFESSLSDPGPGKQ-------------- 522

Qy        158 VGDSEESSILYAPVLQLPLVGEVTFFD----------FQSAERKG---EISQLKSMLTT 203
              || :   |  |       | |  ||            |  |: |   |:    ::|||
Db        523 -DDSGQDVPLRVP-------GVVPVFDAHNLSVGSDYYIQLEEKSGSNLELDYPPALLTT 574

Qy        204 TVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVH 263
                      |:| :|   :| :|:|| | :   |   ||:  | :|     :::|
Db        575 ----------DMDNPERTGPELSQLTALRSVE---LEESSTDEDFFQSSTDPKDSSLPG 620

Qy        264 LSGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQ----TVSQHLH 319
                :   |::    :  :| :|     :||||   |  : : :   ||     |:|  |
Db        621 DLHVTSGPESPFNNIFNDVDKSE----DLPSHQ--KIFDLMELNGVQADFKPATLSSSLD 674

Qy        320 --------------------------AGNLSELGNLNNMNRDLAFHLLREVSDYFRQSE 352
                                       : ||    | : :|        ::|:|: |
Db        675 NPKESVITGHFEKEKPRKIFDSEPLCLSDNLMHQDNFDPLN-------VQELSENF---- 723

Qy        353 PHSPISFLLEKAIRWGYLSLPELLREMMSE-QNGDALSTIFNAAGLNHLD---------- 401
                || ||  :  | ||   | : :: :| :|     :   :  | ||
Db        724 -----LFLQEKNLLKGSLSSKEHINDLQTELKNAGFTEAMLETSCRNSLDTELQFAENKP 778

Qy        402 ---------------QVLL-----------PEVSTPTVGIESPQTPQAKPSVSDPRSV 433
                            |:|           |||  |    |: :||:  |  | |
Db        779 GLSLLQENVSTKGDDTDVMLTGDTLSTSLQSSPEVQVPPTSFETEETPRRVPPDSLPTQG 838

Qy        434 EEHVSQTSPVDTQSKQDQKPQSSATSALS 462
              |   :| : :|    :|   |  :  |::
Db        839 E---TQPTCLDVIVPEDCLHQDISPDAVT 864
```

RESULT 4
US-09-091-501B-8
; Sequence 8, Application US/09091501B
; Patent No. 6518413
; GENERAL INFORMATION:
; APPLICANT: Tinsley, Jonathon M
; APPLICANT: Davies, Kay E
; TITLE OF INVENTION: Utrophin gene expression
; FILE REFERENCE: 620-42
; CURRENT APPLICATION NUMBER: US/09/091,501B
; CURRENT FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: PCT/GB96/03156
; PRIOR FILING DATE: 1996-12-19
; PRIOR APPLICATION NUMBER: GB 9525962.8
; PRIOR FILING DATE: 1995-12-19
; PRIOR APPLICATION NUMBER: GB 9615797.9
; PRIOR FILING DATE: 1996-07-26
; PRIOR APPLICATION NUMBER: GB 9622174.2
; PRIOR FILING DATE: 1996-10-24
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 2008
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (239) ... (250)
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric; Xaa = Unknown
US-09-091-501B-8

Query Match 5.9%; Score 138; DB 4; Length 2008;
Best Local Similarity 21.2%; Pred. No. 0.0023;
Matches 105; Conservative 72; Mismatches 172; Indels 146; Gaps 19;



RESULT 5
US-09-252-991A-17293
; Sequence 17293, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17293
; LENGTH: 357
; TYPE: PRT
; ORGANISM: *Pseudomonas aeruginosa*
US-09-252-991A-17293

Query Match 5.9%; Score 137.5; DB 4; Length 357;
Best Local Similarity 21.6%; Pred. No. 0.00016;
Matches 91; Conservative 67; Mismatches 145; Indels 119; Gaps 18;

RESULT 6
US-09-336-447A-11
; Sequence 11, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIĆ J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024

```
;  CURRENT APPLICATION NUMBER: US/09/336,447A
;  CURRENT FILING DATE:  1999-06-21
;  NUMBER OF SEQ ID NOS: 98
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
;   LENGTH: 610
;   TYPE: PRT
;   ORGANISM: Moraxella catarrhalis
US-09-336-447A-11

  Query Match             5.4%;  Score 126;  DB 4;  Length 610;
  Best Local Similarity   20.8%;  Pred. No. 0.0045;
  Matches   82;  Conservative   71;  Mismatches  166;  Indels   76;  Gaps   15;

Qy          43 QTALRKLSQNPSADERDALQ--EACLNKWKIL--------SDSLYEQFSKTTRDIEL--- 89
               |  :: || |     |: | |:  :  :||: :|        :| |     | |  ||
Db          37 QDSISKLVQ----DDIDTLKQDQQKMNKYLLLNQLANTLITDELNNNVIKNTNSIEALGD 92

Qy          90 -ISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDKGKEREQAD 148
                | |       |: :|    :   | :  |:|                |:   : |||
Db          93 EIGWLENDIADLEEGVEELTKNQNTLIEKDEEH----------------DRLIAQNQAD 135

Qy         149 AK------VKAFFQLVGD--SEESSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSM 200
                :       |:  | | |    :|: |          : |:  || : ||| |||
Db         136 IQTLENNVVEELFNLSGRLIDQEADI----AKNNASIEELYDFDNEVAERIGEIHAYTEE 191

Qy         201 LTTTVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSL----GSQSTNFGFAKSLLTR 256
               :  |        ::  : |: :    :|:  | :     : :       |  :     |:|
Db         192 VNKT-------LENLITNSVKNTDNIDKNKADIDNNINHIYELAQQQDQHSSDIKTLKNN 244

Qy         257 VENALVHLSGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQ 316
               ||  |: |||  :  ||:  | : :  ||:| || |    :  : :::    ::  : :
Db         245 VEEGLLELSGHLIDQKADL-TKDIKALESNVEEGLL--DLSGRLLDQKADLTKDIKALES 301

Qy         317 HLHAGNLSELGNLNNMNRDLA--------FHLLREVSDYFRQSEPHSPISFLLEKAIRWG 368
               ::  |'|    | | :    |:|              |: | : | :  :  :  | ||
Db         302 NVEEGLLDLSGRLLDQKADIAQNQTDIQDLAAYNELQDQYAQKQTEAIDA--LNKASSEN 359

Qy         369 YLSLPEL-----LREMMSEQNGDALSTIFNAAGLN 398
               : : :|     |::  ::|  :|:  :  |:  |
Db         360 TQNIEDLAAYNELQDAYAKQQTEAIDALNKASSEN 394


RESULT 7
US-09-252-991A-30776
; Sequence 30776, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 30776
;   LENGTH: 949
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30776

  Query Match             5.2%;  Score 123;  DB 4;  Length 949;
  Best Local Similarity   23.8%;  Pred. No. 0.018;
  Matches  108;  Conservative   69;  Mismatches  203;  Indels   74;  Gaps   20;

Qy           5 KHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALR--KLSQNPSADERDALQ 62
               | ::|:  | |             |:    |  :   || ||   : :|  :|:| : |
Db         490 KERLEKELKQLKTQQSVAADRAASKAQAEKLYQDVLDAQKALEDFRRTQTLTAEEPEKLA 549

Qy          63 -----EACLNKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWL-AD 116
                    ||  :: |  ||:  |:  : :  ::|::  :      : ||| |    : | ||
Db         550 ELAQLEASQDELKRSSDAFAERVQQLSARLQLVARQLGDLEAKERTLEDALRRRQLLPAD 609

Qy         117 L------SEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAP 170
               |      |   | |: :||:  |    |  :  :: | :::|            |||
Db         610 LPFGTPFMEPVDDSLDNLLPL--LNDYQDTWQALQRIDGQIEA------------LYA- 653

Qy         171 VLQLPLVGEVTFFDFQSAERKGEISQLKSMLTTTVAQERFAIQFKMENAKR-CVTQLDRL 229
                 |: | |   |       ||:     || ::     ::  |:   :  |:|  || : |
Db         654 --QVRLKGVAKFDSEDDTERR-----LKLLINAYAHRQDEAL--TLAKARRAAVTDIAR- 703

Qy         230 SALVSTKCHSLGSQSTNFGFAKSLLTRVEN----ALVHLSGIKLAPKAEA-KTVEQEV-A 283
                     :: |   |      :|  | |    : :    | |||  :| | ::| : :
Db         704 ------TLRNIRSDYDNLEHQLALFNREINKRQVSNLQSFRIVLAPNKDALKHIDQIIHS 757

Qy         284 ESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRDLAFHLLRE 343
                   |||  |  |           :|:|:     | | | ::||       :|| | |  |
Db         758 AGQYEEGENLSVFDLTQSAEQDAKNEEAKEYLARLVAANHNQLG-----LKDL-FELAFE 811

Qy         344 VSDYFRQSEPHSPI-------SFLLEKAIRWGYLSLPELLREMMSEQNGDALSTIF--NA 394
               ::    |   |: |       : :  ||:   ||    ||  |  || |      :   |
Db         812 ITKVNGQPIMHTDIDGAASNGTTMTIKALTNMYL----LLHLMDREQAGRVRLPYYLDEA 867

Qy         395 AGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVS 428
               | ::  :|  | | |     :  |     || ||
Db         868 ADIDERNQAALLETSLQLGFV--PILASVKPQVS 899


RESULT 8
US-08-770-301A-1
; Sequence 1, Application US/08770301A
; Patent No. 5948637
;  GENERAL INFORMATION:
;    APPLICANT:  IKEDA, JUN
;    APPLICANT:  KANEDA, SUMIKO
;    APPLICANT:  YANAGI, HIDEKI
;    APPLICANT:  MATSUMOTO, MASAYASU
;    APPLICANT:  YURA, TAKASHI
;    TITLE OF INVENTION:  NOVEL STRESS PROTEINS
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH AND BIRCH
;      STREET:  PO BOX 747
;      CITY:  FALLS CHURCH
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22040-0747
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,301A
;      FILING DATE:  20-DEC-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURPHY JR, GERALD M
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1422-287
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)-205-8000
;      TELEFAX:  (703)-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  999 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-770-301A-1

  Query Match             5.1%;  Score 120.5;  DB 2;  Length 999;
```

```
  Best Local Similarity   22.2%;  Pred. No. 0.033;
  Matches   89;  Conservative   63;  Mismatches  122;  Indels  127;  Gaps   19;

Qy         81 SKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDK 140
              ::  | :| |   :    | |   :  | |:: | ||               ||:  | :
Db        690 ARKRRMVEEIGVELVVLDLPDLPEDKLAQSVQKLQDL--------------TLR--DLE 732

Qy        141 GKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSM 200
              :|||:|   ::||       |    || | |     ||:  : |  |  |::|   :
Db        733 KQEREKAANSLEAFI-----FETQDKLYQPEYQ-----EVS-TEEQREEISGKLSAASTW 781

Qy        201 L-------TTTVAQERFA--------IQFKMENAKRCVTQLDRLSALVSTKCHSLGSQST 245
              |       || : :|: |        : |::|  |:     :|||| :   ||
Db        782 LEDEGVGATTVMLKEKLAELRKLCQGLFFRVEERKKWP---ERLSALDNLLNHS------ 832

Qy        246 NFGFAKSLLTRVENALVHLSGIKLAPKAEA--KTVEQEVAESSVSEGELPSHMDTKHIER 303
                  : | | :| |: :     ||    |  ::|     :        :
Db        833 --------------SMFLKGARLIPEMDQIFTEVEMTTLEKVINETWAWKNATLAEQAK 877

Qy        304 IPMASEQAQTVSQHLHAGNLSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEK 363
              :| |:|:   :|: : |  ::              | |||             :|| |
Db        878 LP-ATEKPVLLSKDIEAKMMA--------------LDREV------------QYLLNK 908

Qy        364 AIRWGYLSLPELLREMMSEQNGDALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQA 423
              |      : |   |    ::||       ||:  :  ::|: |          : || |
Db        909 A----KFTKP---RPRPKDKNGTRAEPPLNASASDQGEKVIPP----------AGQTEDA 951

Qy        424 KPSVSDPRSVEEHVSQTSPVDTQ--------SKQDQKPQSS 456
              :| :|:|  ||   : : | ||:        :: :|| ||:
Db        952 EP-ISEPEKVE---TGSEPGDTEPLELGGPGAEPEQKEQST 988


RESULT 9
US-09-175-581-1
; Sequence 1, Application US/09175581
; Patent No. 6034232
;  GENERAL INFORMATION:
;    APPLICANT:  IKEDA, JUN
;    APPLICANT:  KANEDA, SUMIKO
;    APPLICANT:  YANAGI, HIDEKI
;    APPLICANT:  MATSUMOTO, MASAYASU
;    APPLICANT:  YURA, TAKASHI
;    TITLE OF INVENTION:  NOVEL STRESS PROTEINS
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH AND BIRCH
;      STREET:  PO BOX 747
;      CITY:  FALLS CHURCH
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22040-0747
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/175,581
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/770,301
;      FILING DATE:  20-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURPHY JR, GERALD M
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1422-287
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)-205-8000
;      TELEFAX:  (703)-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  999 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-175-581-1

  Query Match             5.1%;  Score 120.5;  DB 3;  Length 999;
  Best Local Similarity   22.2%;  Pred. No. 0.033;
  Matches   89;  Conservative   63;  Mismatches  122;  Indels  127;  Gaps   19;

Qy         81 SKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDK 140
              ::  | :| |   :    | |   :  | |:: | ||               ||:  | :
Db        690 ARKRRMVEEIGVELVVLDLPDLPEDKLAQSVQKLQDL--------------TLR--DLE 732

Qy        141 GKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSM 200
              :|||:|   ::||       |    || | |     ||:  : |  |  |::|   :
Db        733 KQEREKAANSLEAFI-----FETQDKLYQPEYQ-----EVS-TEEQREEISGKLSAASTW 781

Qy        201 L-------TTTVAQERFA--------IQFKMENAKRCVTQLDRLSALVSTKCHSLGSQST 245
              |       || : :|: |        : |::|  |:     :|||| :   ||
Db        782 LEDEGVGATTVMLKEKLAELRKLCQGLFFRVEERKKWP---ERLSALDNLLNHS------ 832

Qy        246 NFGFAKSLLTRVENALVHLSGIKLAPKAEA--KTVEQEVAESSVSEGELPSHMDTKHIER 303
                  : | | :| |: :     ||    |  ::|     :        :
Db        833 --------------SMFLKGARLIPEMDQIFTEVEMTTLEKVINETWAWKNATLAEQAK 877

Qy        304 IPMASEQAQTVSQHLHAGNLSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEK 363
              :| |:|:   :|: : |  ::              | |||             :|| |
Db        878 LP-ATEKPVLLSKDIEAKMMA--------------LDREV------------QYLLNK 908

Qy        364 AIRWGYLSLPELLREMMSEQNGDALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQA 423
              |      : |   |    ::||       ||:  :  ::|: |          : || |
Db        909 A----KFTKP---RPRPKDKNGTRAEPPLNASASDQGEKVIPP----------AGQTEDA 951

Qy        424 KPSVSDPRSVEEHVSQTSPVDTQ--------SKQDQKPQSS 456
              :| :|:|  ||   : : | ||:        :: :|| ||:
Db        952 EP-ISEPEKVE---TGSEPGDTEPLELGGPGAEPEQKEQST 988


RESULT 10
US-09-541-782-8
; Sequence 8, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
;  APPLICANT:  Nislow, Corey
;  APPLICANT:  Sakowicz, Roman
;  APPLICANT:  Beraud, Christophe
;  TITLE OF INVENTION:  Antifungal Assay
;  FILE REFERENCE:  1015
;  CURRENT APPLICATION NUMBER: US/09/541,782
;  CURRENT FILING DATE:  2000-04-03
;  NUMBER OF SEQ ID NOS:  10
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 1066
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-541-782-8

  Query Match             5.1%;  Score 119.5;  DB 3;  Length 1066;
  Best Local Similarity   19.1%;  Pred. No. 0.046;
  Matches   96;  Conservative   78;  Mismatches  161;  Indels  167;  Gaps   21;

Qy          5 KHQIEQLSKPLSDDSIC-----GVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADER- 58
              |   |:: | | | :       |:||          |    :  |:  |||   :|:
Db        373 KEYTEEIDK-LKRDLMAARDKNGIYLA----------EETYGEITLKLESQNRELNEKM 420

Qy         59 ---DALQEACLNKWKILSD---SLYEQ----------------------FSKTTRDIE 88
                 ||::.  || || |:   || |:                      :|| | :
Db        421 LLLKALKDELQNKEKIFSEVSMSLVEKTQELKKTEENLLNTKGTLLLTKKVLTKTKRRYK 480

Qy         89 LISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNPVLP-----VETLKSDDDKGKE 143
                 ||:    :  | : |  :   |||:       |:  :        | ::   |: |:
Db        481 EKKELVASHMKTEQVLTTQAQEILAAADLATDDTHQLHGTIERRRELDEKIRRSCDQFKD 540
```

```
Qy        144 REQADAKVKAFFQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSMLTT 203
Db        541 RMQDN------LEMIG-----------------GSLNLYQDQQAALKEQLSQ--EMVNS 574

Qy        204 TVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVH 263
Db        575 SYVSQRLAL-----NSSKSIEMLKEMCA------QSLQDQ-----------TNLHNKLI- 611

Qy        264 LSGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNL 323
Db        612 ---------GEVMKISDQHSQAFVA--KLMEQMQQQQL----LMSKEIQTNLQVIEENNQ 656

Qy        324 SELGNLNNMNRDLAFHL---LREVSDYFRQSEPHSPISFLLEKAIRWGYLSLP------- 373
Db        657 RHKAMLDSMQEKFATIIDSSLQSVEEHAKQ---------MHKKLEQLGAMSLPDAEELQN 707

Qy        374 ---ELLREMMSEQNGDAL----------------STIFNAAGLNHLDQVLLPE------ 407
Db        708 LQEELANERALAQQEDALLESMMMQMEQIKNLRSKNSISMSVHLNKMEESRLTRNHRIDD 767

Qy        408 -----VSTPTVGIESPQTPQAK 424
Db        768 IKSGIQDYQKLGIEASQSAQAE 789
```

RESULT 11
US-09-723-820-8
; Sequence 8, Application US/09723820
; Patent No. 6468760
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/723,820
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-723-820-8

Query Match 5.1%; Score 119.5; DB 4; Length 1066;
Best Local Similarity 19.1%; Pred. No. 0.046;
Matches 96; Conservative 78; Mismatches 161; Indels 167; Gaps 21;

```
Qy          5 KHQIEQLSKPLSDDSIC-----GVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADER- 58
Db        373 KEYTEEIDK-LKRDLMAARDKNGIYLA-----------EETYGEITLKLESQNRELNEKM 420

Qy         59 ---DALQEACLNKWKILSD---SLYEQ-----------------------FSKTTRDIE 88
Db        421 LLLKALKDELQNKEKIFSEVSMSLVEKTQELKKTEENLLNTKGTLLLTKKVLTKTKRRYK 480

Qy         89 LISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNPVLP-----VETLKSDDDKGKE 143
Db        481 EKKELVASHMKTEQVLTTQAQEILAAADLATDDTHQLHGTIERRRELDEKIRRSCDQFKD 540

Qy        144 REQADAKVKAFFQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSMLTT 203
Db        541 RMQDN------LEMIG-----------------GSLNLYQDQQAALKEQLSQ--EMVNS 574

Qy        204 TVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVH 263
Db        575 SYVSQRLAL-----NSSKSIEMLKEMCA------QSLQDQ-----------TNLHNKLI- 611

Qy        264 LSGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNL 323
Db        612 ---------GEVMKISDQHSQAFVA--KLMEQMQQQQL----LMSKEIQTNLQVIEENNQ 656

Qy        324 SELGNLNNMNRDLAFHL---LREVSDYFRQSEPHSPISFLLEKAIRWGYLSLP------- 373
Db        657 RHKAMLDSMQEKFATIIDSSLQSVEEHAKQ---------MHKKLEQLGAMSLPDAEELQN 707

Qy        374 ---ELLREMMSEQNGDAL----------------STIFNAAGLNHLDQVLLPE------ 407
Db        708 LQEELANERALAQQEDALLESMMMQMEQIKNLRSKNSISMSVHLNKMEESRLTRNHRIDD 767

Qy        408 -----VSTPTVGIESPQTPQAK 424
Db        768 IKSGIQDYQKLGIEASQSAQAE 789
```

RESULT 12
US-10-270-085-8
; Sequence 8, Application US/10270085
; Patent No. 6627408
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/10/270,085
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/09/723,820
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-270-085-8

Query Match 5.1%; Score 119.5; DB 4; Length 1066;
Best Local Similarity 19.1%; Pred. No. 0.046;
Matches 96; Conservative 78; Mismatches 161; Indels 167; Gaps 21;

```
Qy          5 KHQIEQLSKPLSDDSIC-----GVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADER- 58
Db        373 KEYTEEIDK-LKRDLMAARDKNGIYLA-----------EETYGEITLKLESQNRELNEKM 420

Qy         59 ---DALQEACLNKWKILSD---SLYEQ-----------------------FSKTTRDIE 88
Db        421 LLLKALKDELQNKEKIFSEVSMSLVEKTQELKKTEENLLNTKGTLLLTKKVLTKTKRRYK 480

Qy         89 LISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNPVLP-----VETLKSDDDKGKE 143
Db        481 EKKELVASHMKTEQVLTTQAQEILAAADLATDDTHQLHGTIERRRELDEKIRRSCDQFKD 540

Qy        144 REQADAKVKAFFQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSMLTT 203
Db        541 RMQDN------LEMIG-----------------GSLNLYQDQQAALKEQLSQ--EMVNS 574

Qy        204 TVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVH 263
Db        575 SYVSQRLAL-----NSSKSIEMLKEMCA------QSLQDQ-----------TNLHNKLI- 611

Qy        264 LSGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNL 323
Db        612 ---------GEVMKISDQHSQAFVA--KLMEQMQQQQL----LMSKEIQTNLQVIEENNQ 656

Qy        324 SELGNLNNMNRDLAFHL---LREVSDYFRQSEPHSPISFLLEKAIRWGYLSLP------- 373
Db        657 RHKAMLDSMQEKFATIIDSSLQSVEEHAKQ---------MHKKLEQLGAMSLPDAEELQN 707

Qy        374 ---ELLREMMSEQNGDAL----------------STIFNAAGLNHLDQVLLPE------ 407
```

```
                ||  |    |  |||                  ::|  :  || :::  |
Db        708 LQEELANERALAQQEDALLESMMMQMEQIKNLRSKNSISMSVHLNKMEESRLTRNHRIDD 767

Qy        408 -----VSTPTVGIESPQTPQAK 424
                   :|||: |: ||:
Db        768 IKSGIQDYQKLGIEASQSAQAE 789


RESULT 13
US-09-976-594-321
; Sequence 321, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
;  APPLICANT: Furness, Michael
;  APPLICANT:  Buchbinder, Jenny
;  TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
;  FILE REFERENCE: PA-0041 US
;  CURRENT APPLICATION NUMBER: US/09/976,594
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: 60/240,409
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 1143
;  SOFTWARE: PERL Program
; SEQ ID NO 321
;   LENGTH: 806
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6673549 3768043CD1
US-09-976-594-321

  Query Match             5.1%;  Score 118.5;  DB 4;  Length 806;
  Best Local Similarity   20.7%;  Pred. No. 0.036;
  Matches  110;  Conservative   80;  Mismatches  217;  Indels  125;  Gaps   20;

Qy          4 SKH--------QIEQLSKPLSDDS------------ICGVYLKLEKS---AFRPLRNEFN 40
              |||          | :  :  | |:|           ||  :    |:    | :   || :
Db        158 SKHSQQDELMADISKSIQSLEDESKSLKSQIAEAKIICKTFKMSEERRAIAIKDALNENS 217

Qy         41 VAQTALRKLSQNPSADERDALQEACLNKWKILSDSLYEQFSKTTRDIELISWFVAAQFLL 100
                ||: ::| |  :   :  : |  ||| ||
Db        218 QLQTSHKQLFQQEAEVWKGEVSE--LNKQKI---------------------------- 246

Qy        101 DTTLESAANSLEWLADLSEKHWDHLNPVLP--------VETLKSDDDKGKEREQADAKVK 152
                | | :    | : :  | |   |   ||         :|   :|||  :    ::::
Db        247 --TFEDSKVHAEQVLNDKENHIKTLTGHLPMMKDQAAVLEEDTTDDDNLELEVNSESENG 304

Qy        153 AFFQLVGDSEESSILYAPVLQL---PLVGEVTFFDFQSAE---RKGEISQ-LKSMLTTTV 205
                |:          :::|  |       | ||       | :|    | |::: :|:: |
Db        305 AYLDNPPKGALKKLIHAAKLNASLKTLEGERNQIYIQLSEVDKTKEELTEHIKNLQTQQA 364

Qy        206 AQERFAIQFKMENAK-----RCVTQLDRLSALVSTKCHSLGSQSTNFGFAK-SLLTRVEN 259
              : :     |: || |     : :|:| : : :   | |   :   |:   |   |::|:
Db        365 SLQSENTHFENENQKLQQKLKVMTELYQENEM---KLHRKLTVEENYRLEKEEKLSKVDE 421

Qy        260 ALVHLSGIKLAPKAEAKTVEQEVAESSVS-EGELPSHMDTKHIERIPMASEQAQTVSQHL 318
              : | :      :  || :|:|:  :  | :|:: ||    |     :|:  |:
Db        422 KISHATEELETYRKRAKDLEEELERTIHSYQGQIISHEKKAHDNW--LAARNAER----- 474

Qy        319 HAGNLSELGNLNNMNR------DLAFHLLRE--------VSDYFRQSEPHSPISFLLEKA 364
                ||::|   |  ||      :| | ||:         : : |:  |: |       :
Db        475 ---NLNDLRKENAHNRQKLTETELKFELLEKDPYALDVPNTAFGREHSPYGPSPLGWPSS 531

Qy        365 IRWGYLSLPELLREMM-------------SEQNGDALS-TIFNAAGLNHLDQVLLPEVST 410
              :|| | ||   :                |   |: |   | |  | :  |::  |  :
Db        532 ETRAFLSPPTLLEGPLRLSPLLPGGGGRGSRGPGNPLDHQITNERGESSCDRLTDPHRAP 591

Qy        411 PTVGIESPQTPQAKPSVSDPRSVEEHVSQTSPVDTQSKQDQKPQSSATSALS 462
              .  |  ||      |   | | :    |: |       | |    | :  ||
Db        592 SDTGSLSP------PWDQDRRMMFPPPGQSYPDSALPPQRQDRFCSNSGRLS 637


RESULT 14
US-09-976-594-726
; Sequence 726, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
;  APPLICANT: Furness, Michael
;  APPLICANT:  Buchbinder, Jenny
;  TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
;  FILE REFERENCE: PA-0041 US
;  CURRENT APPLICATION NUMBER: US/09/976,594
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: 60/240,409
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 1143
;  SOFTWARE: PERL Program
; SEQ ID NO 726
;   LENGTH: 2468
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1
US-09-976-594-726

  Query Match             4.9%;  Score 115.5;  DB 4;  Length 2468;
  Best Local Similarity   21.4%;  Pred. No. 0.42;
  Matches  119;  Conservative   78;  Mismatches  199;  Indels  161;  Gaps   27;

Qy          2 PLSKHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQN--------- 52
              | |  : :: |:    |: |     |  |   |  : :::  :  |:| :
Db       1209 PPSSMEEDKFSRSALRDAYCS-----EVKASTTLDIKDSISAVSSEKVSPSKSPSLSPSP 1263

Qy         53 PSADERDALQEACLNKWKILSDSLYEQFSKTTRDIEL--------ISWFVAAQFL----- 99
              ||  |:  | |  :|           || |  :|::        :|  |  : :
Db       1264 PSPLEKTPLGERSVN-----------FSLTPNEIKVSAEAEVAPVSPEVTQEVVEEHCA 1311

Qy         100 --LDTTLE----------SAANSLEWLADLSEKHWDHL-------NPVLPVETLKSDDDK 140
                 | |||          || ::  : :   ||  ||        | :||    ||
Db       1312 SPEDKTLEVVSPSQSVTGSAGHTPYYQSPTDEKS-SHLPTEVIEKPPAVPVSFEFSDAKD 1370

Qy         141 GKEREQADAKVKAFFQLVGDSEES-SILYAPVLQLPLVG-EVTFFDFQSAERKG------ 192
                ||    | |    : | |||     : :|:   ||:| |  :  | ||: |
Db       1371 ENER----ASVSPMDEPVPDSESPIEKVLSPLRSPPLIGSESAYESFLSADDKASGRGAE 1426

Qy         193 --------------EISQLKSMLTTTVAQER-------FA---IQFKMENAKRCVTQLDR 228
                           ::| :  | :|:: |::       ||     |  |      |  :
Db       1427 SPFEEKSGKQGSPDQVSPVSEMTSTSLYQDKQEGKSTDFAPIKEDFGQEKKTDDVEAMSS 1486

Qy         229 LSALVSTKCHSLGSQS------TNFGFAKSLLTRVENALVHLSGIKLAPKAEAKTVEQEV 282
                ||   :  ||  |       : ||  |       |:  : :|   ::  |: |  |:: |
Db       1487 QPALALDE-RKLGDVSPTQIDVSQFGSFK------EDTKMSISEGTVSDKS-ATPVDEGV 1538

Qy         283 AESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRDLAFHLLR 342
               || : |  | : : |  :                ||  |||    |:|: ::
Db       1539 AEDTYSHMEGVASVSTASVATSSFPEPTTDDVSPSLHA----EVGSPHS----------T 1584

Qy         343 EVSDYFRQSEPHSPISF-------LLEKAIRWGYLSLPEL----------LREMMSEQNG 385
               || |    |   :| :|          |:  |   :| |:          ::: |||:
Db       1585 EVDDSLSVSVVQTPTTFQETEMSPSKEECPRPMSISPPDFSPKTAKSRTPVQDHRSEQS- 1643

Qy         386 DALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPRSVEEHVSQTSPVDT 445
                ::|  |   |    :| |  : |       :||  |      |        |:::  | :
Db       1644 -SMSIEF---GQESPEQSLAMDFSR-----QSPDHPTVGAGVL-------HITENGPTEV 1687

Qy         446 QSKQDQKPQSSATSALS 462
                  |  |     |:||
Db       1688 ----DYSPSDMQDSSLS 1700


RESULT 15
US-09-134-000C-5352
; Sequence 5352, Application US/09134000C
; Patent No. 6617156
```

```
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 032796-032
;  CURRENT APPLICATION NUMBER: US/09/134,000C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/055,778
;  PRIOR FILING DATE: 1997-08-15
;  NUMBER OF SEQ ID NOS: 6812
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5352
;   LENGTH: 431
;   TYPE: PRT
;   ORGANISM: Enterococcus faecalis
US-09-134-000C-5352

  Query Match             4.9%;  Score 114.5;  DB 4;  Length 431;
  Best Local Similarity   20.2%;  Pred. No. 0.031;
  Matches   78;  Conservative   69;  Mismatches  137;  Indels  103;  Gaps   16;

Qy          3 LSKHQIEQLSKPLSDDSICGVYLKLEKSAF--------------RPLRNEFNVA------ 42
              ::| :: :  | :|||:        |||  |                :| : |  :|
Db          1 MTKRRLIRKIKDISDDT-------LEKEVFFQPGFEPEKIISGIKPEQQEKEMATLRERL 53

Qy         43 ------QTALRKLSQ------NPSADERDALQEACLNKWKILSDSLYEQFSKTTRDIELI 90
                  :|| :|| :       |     |: ||:|       : |:   :| | :    :|
Db         54 TYYQKKETAFKKLQEENEQLKNQQETEKQALEEKNQALLETLNQKEHEIF-ELEEQLETN 112

Qy         91 SWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNPVLPVETLKSD-DDKGKEREQADA 149
              |         | | |::|    |  : :                :| |:|:   :  | |:  |
Db        113 SETTTEATTLQTNLQNAEKQLTTVHN-------------QMERLESELTQQVAELEKESA 159

Qy        150 KVKAFFQLVGD--SEESSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSM---LTTT 204
              : :    | : :  :||    |                        |:||:|: ::      |||
Db        160 EQQQHIQTLHNQLAEEKGKL--------------------EKKGDITGMEKQVETLTTE 198

Qy        205 VAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVHL 264
              : :     |   |   ::   || :: |    |    ::  :          |  :  ::| | :
Db        199 LEESTTTQQALQEQIEKLTVQLQKVEAEKQTLQEAVQQEV----IVKQRMEELKNELQN- 253

Qy        265 SGIKLAPKAEAKTV--------EQEVAESSVSEGELPS---HMDTK--HIERIPMASEQA 311
                |   | ||:          :: :||  | |||| :    :: |  ::|::    :
Db        254 ---SLQQKNEAEVKLGRLQGVNDKHLAEKQVYEGELAALRKEVEAKEDNLEQVQKELKTL 310

Qy        312 QTVSQHLHAGN--LSELGNLNNMNRDL 336
              | ||:::   :  |:||        ||
Db        311 QAVSKNIEPSSEQLTELLRAKQQINDL 337


Search completed: July  3, 2004, 02:47:19
Job time : 26 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 3, 2004, 02:46:20 ; Search time 81 Seconds
(without alignments)
1779.314 Million cell updates/sec

Title: US-10-780-347-2
Perfect score: 2343
Sequence: 1 MPLSKHQIEQLSKPLSDDSI.....DTQSKQDQKPQSSATSALSW 463

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	143	6.1	884	14	US-10-179-766-10	Sequence 10, Appl
2	134.5	5.7	1179	9	US-09-815-242-13262	Sequence 13262, A
3	134.5	5.7	1179	12	US-10-282-122A-74007	Sequence 74007, A
4	133	5.7	2295	16	US-10-437-963-135452	Sequence 135452,
5	129.5	5.5	860	12	US-10-072-012-838	Sequence 838, App
6	129.5	5.5	860	12	US-10-037-417-59	Sequence 59, Appl
7	129.5	5.5	860	15	US-10-080-334-166	Sequence 166, App
8	127	5.4	1294	12	US-10-282-122A-61292	Sequence 61292, A
9	126.5	5.4	752	14	US-10-171-311-48	Sequence 48, Appl
10	126.5	5.4	785	14	US-10-309-851-24	Sequence 24, Appl
11	126.5	5.4	824	15	US-10-108-260A-3569	Sequence 3569, Ap
12	126.5	5.4	1133	14	US-10-309-851-14	Sequence 14, Appl
13	126.5	5.4	1133	14	US-10-309-851-16	Sequence 16, Appl
14	126.5	5.4	1135	14	US-10-309-851-12	Sequence 12, Appl
15	126	5.4	506	14	US-10-157-223-7	Sequence 7, Appli
16	126	5.4	506	15	US-10-369-493-1648	Sequence 1648, Ap
17	126	5.4	610	10	US-09-952-267-11	Sequence 11, Appl
18	125	5.3	1190	15	US-10-369-493-2128	Sequence 2128, Ap
19	124.5	5.3	1940	10	US-09-738-630-99	Sequence 99, Appl
20	124.5	5.3	1940	16	US-10-408-765A-1175	Sequence 1175, Ap
21	123	5.2	1137	12	US-10-336-472-16	Sequence 16, Appl
22	123	5.2	1524	12	US-10-336-472-18	Sequence 18, Appl
23	123	5.2	1935	12	US-10-336-472-22	Sequence 22, Appl
24	122	5.2	2109	15	US-10-369-493-6346	Sequence 6346, Ap
25	121.5	5.2	3225	16	US-10-408-765A-254	Sequence 254, App
26	121	5.2	715	16	US-10-408-765A-2096	Sequence 2096, Ap
27	121	5.2	1221	12	US-10-282-122A-48350	Sequence 48350, A
28	120	5.1	654	16	US-10-437-963-113490	Sequence 113490,
29	119.5	5.1	1179	9	US-09-815-242-13608	Sequence 13608, A
30	119.5	5.1	2457	16	US-10-437-963-115622	Sequence 115622,
31	119	5.1	1023	12	US-10-282-122A-60535	Sequence 60535, A
32	119	5.1	1557	15	US-10-369-493-2224	Sequence 2224, Ap
33	118.5	5.1	1665	14	US-10-032-585-7470	Sequence 7470, Ap
34	118.5	5.1	2271	12	US-10-282-122A-43924	Sequence 43924, A
35	118	5.0	1979	14	US-10-205-823-419	Sequence 419, App
36	117.5	5.0	651	12	US-10-282-122A-46243	Sequence 46243, A
37	117.5	5.0	2283	14	US-10-172-502-4	Sequence 4, Appli
38	117	5.0	859	16	US-10-437-963-169313	Sequence 169313,
39	117	5.0	1679	15	US-10-369-493-22080	Sequence 22080, A
40	116.5	5.0	1585	16	US-10-437-963-124349	Sequence 124349,
41	116	5.0	620	16	US-10-437-963-103451	Sequence 103451,
42	116	5.0	1029	12	US-10-282-122A-45920	Sequence 45920, A
43	116	5.0	1099	16	US-10-437-963-177644	Sequence 177644,
44	116	5.0	1583	16	US-10-408-765A-1635	Sequence 1635, Ap
45	116	5.0	1616	10	US-09-820-843A-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-10-179-766-10
; Sequence 10, Application US/10179766
; Publication No. US20030190312A1
; GENERAL INFORMATION:
; APPLICANT: Kenyon, Cynthia
; APPLICANT: Apfeld, Javier
; APPLICANT: Dillin, Andrew
; APPLICANT: Garigan, Delia
; APPLICANT: Hsu, Ao-Lin A.
; APPLICANT: Lehrer-Graiwer, Josh
; APPLICANT: Murphy, Coleen
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Eukaryotic Genes Involved in Adult Lifespan Regulation
; FILE REFERENCE: 023070-119930US
; CURRENT APPLICATION NUMBER: US/10/179,766
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: US 60/300,577
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/301,052
; PRIOR FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/373,975
; PRIOR FILING DATE: 2002-04-18
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 884
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
; OTHER INFORMATION: llw-4 (long-lived worm protein), gene F45H10.4
US-10-179-766-10

Query Match 6.1%; Score 143; DB 14; Length 884;
Best Local Similarity 20.7%; Pred. No. 0.0021;
Matches 103; Conservative 99; Mismatches 165; Indels 130; Gaps 25;

Qy 27 LEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACLNKW----KILSDSLYEQFSK 82

```
Db        218 LRPDAYTEARITRRLALTSTATLSQR---------RERFLKKYSQIRQVVAQNLFGQQQR 268

Qy         83 TTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDKGK 142

Db        269 LTRPVEDLVTETSSKLAVEEAPEETTTQEETTTDASE----------VTTTKAVEEATE 317

Qy        143 E-REQADAKVKAFFQLVGDSEESSILYAPVLQ-------------LPLVGEVTFFD-FQ 186

Db        318 EVTEEATEATEA---PVATTKESSEMHVNTIRNMIRSASEKDLSKYVTLISEGKFSELFE 374

Qy        187 SAERK------------GEISQLKSMLTTTVAQ-------ERFAIQFKMENAKRCVTQL 226

Db        375 LAEQKKLTLTSKFDEKLSSKMAKLKDLINEALSEKEKSGEIEQAMEKFEKPEKSELVAME 434

Qy        227 DRLSALVSTKCHSLGSQSTNFGFAKSLLTR----VENALVHLSGIKLAPKAEAKTVEQE- 281

Db        435 DKDTPAVFTISDSLKHKKAEAKLAHTIVSRNVVEAENA---IEKEVVEPKAEEKKVKEED 491

Qy        282 ---VAESSVSE---GELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNM--N 333

Db        492 VKAVAEEKKEEKKPGKLP--MKIEKLEK-PVDTK-------------SENHELKKVLDD 534

Qy        334 RDLAFHLLREVSDYFRQSEPHSPISFLLEK---AIRWGYLSLPELLREMMSEQ---NGDA 387

Db        535 KERALLVESEIKNTAEETKPKVE-SFKSEETTVAID----DMPALEKEESAEKKETTGEP 589

Qy        388 LSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPRSVEEHVSQTSPVDTQ- 446

Db        590 TTT---------------EAAVETTEASETPKPEAKPELLS--NLEDVLTLTTP-ETET 630

Qy        447 ---SKQDQKPQSSATSA 460

Db        631 IEGSGEREEPTTSAPAA 647


RESULT 2
US-09-815-242-13262
; Sequence 13262, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13262
;   LENGTH: 1179
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
US-09-815-242-13262

  Query Match             5.7%;  Score 134.5;  DB 9;  Length 1179;
  Best Local Similarity   21.6%;  Pred. No. 0.019;
  Matches  100;  Conservative   80;  Mismatches  178;  Indels  105;  Gaps   24;

Qy          5 KHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEA 64

Db        674 KPELEQLQKEIAADE---ASLGSEEAALKTLQDQMAALTERLEAIK---SQGEQARIQEQ 727

Qy         65 CLNKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDH 124

Db        728 GLS-------LAYQQTSQQVEELETL-WKLQEE-EIDRLSEG-----DWQAD-KEKCQES 772

Qy        125 L--------NPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPL 176

Db        773 LATIASEKQNLEAEIEEIKSNKNAIQERYQ-------------NLQEEVAQARLLKTEL 818

Qy        177 VGEVTFFDFQSAERKGEISQLKSMLTTTVAQERF--AIQFKMENAKRCVTQLDRLSALVS 234

Db        819 QGQKR-YEVADIERLG-----KELDNLNIEQEEIQRMLQEKVDNLEKVDTE------LLS 866

Qy        235 TKCHSLGSQSTNF--GFAKSL--LTRVENALVHLSG-----------IKLAPKAEAKTV 278

Db        867 QQAEESKTQKTNLQQGLIRKQFELDDIEGQLDDIASHLDQARQQNEEWIRKQTRAEAK-- 924

Qy        279 EQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLS--ELGNLNNMNRDL 336

Db        925 KEKVSERL-------RHLQNQLTDQYQISYTEALEKAHELENLNLAEQEVQDLEKAIRSL 977

Qy        337 AFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDA---LSTIFN 393

Db        978 GPVNLEAIDQY---EEVHNRLDFL--NSQRDDILSAKNLLLETITEMNDEVKERFKSTFE 1032

Qy        394 AA------------GLNHLDQVLLPEVSTPTVGIESPQTPQAK 424

Db       1033 AIRESFKVTFKQMPGGGQAD-LILTEGDLLTAGVEISVQPPGK 1074


RESULT 3
US-10-282-122A-74007
; Sequence 74007, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
```

; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 74007
; LENGTH: 1179
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-74007

Query Match 5.7%; Score 134.5; DB 12; Length 1179;
Best Local Similarity 21.6%; Pred. No. 0.019;
Matches 100; Conservative 80; Mismatches 178; Indels 105; Gaps 24;

```
Qy          5 KHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEA 64
              | ::||| | :: |       |  |::| : |:::     |  :    |  |:  :||
Db        674 KPELEQLQKEIAADE---ASLGSEEAALKTLQDQMAALTERLEAIK---SQGEQARIQEQ 727

Qy         65 CLNKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDH 124
               |:        |:| |:   ::| : | :  :  :|    |      :| ||  ||  :
Db        728 GLS-------LAYQQTSQQVEELETL-WKLQEE-EIDRLSEG-----DWQAD-KEKCQES 772

Qy        125 L--------NPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPL 176
              |        |    :| :||: :  :|| |               : :  :  | :|:  |
Db        773 LATIASEKQNLEAEIEEIKSNKNAIQERYQ-------------NLQEEVAQARLLKTEL 818

Qy        177 VGEVTFFDFQSAERKGEISQLKSMLTTTVAQERF--AIQFKMENAKRCVTQLDRLSALVS 234
              |:  ::   || |    | :   : ||     :| |::| ::  |:       |:|
Db        819 QGQKR-YEVADIERLG-----KELDNLNIEQEEIQRMLQEKVDNLEKVDTE------LLS 866

Qy        235 TKCHSLGSQSTNF--GFAKSL--LTRVENALVHLSG-----------IKLAPKAEAKTV 278
                :    :| ||   |  :   |  :|  |  ::             |:   :||||
Db        867 QQAEESKTQKTNLQQGLIRKQFELDDIEGQLDDIASHLDQARQQNEEWIRKQTRAEAK-- 924

Qy        279 EQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLS--ELGNLNNMNRDL 336
              :::|:|          |:  :  ::  ::  :|   :  |   ||:  |: :|   | |
Db        925 KEKVSERL-------RHLQNQLTDQYQISYTEALEKAHELENLNLAEQEVQDLEKAIRSL 977

Qy        337 AFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDA---LSTIFN 393
              |  :  |    | |: : ||   :  |   ||  || | ::| | :      : |
Db        978 GPVNLEAIDQY---EEVHNRLDFL--NSQRDDILSAKNLLLETITEMNDEVKERFKSTFE 1032

Qy        394 AA-----------GLNHLDQVLLPEVSTPTVGIESPQTPQAK 424
              |           |    | ::| |    | |:|    | |
Db       1033 AIRESFKVTFKQMFGGGQAD-LILTEGDLLTAGVEISVQPPGK 1074
```

RESULT 4
US-10-437-963-135452
; Sequence 135452, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 135452
; LENGTH: 2295
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_37127C.1.pep
US-10-437-963-135452

Query Match 5.7%; Score 133; DB 16; Length 2295;
Best Local Similarity 20.7%; Pred. No. 0.07;
Matches 118; Conservative 68; Mismatches 176; Indels 208; Gaps 25;

```
Qy          4 SKHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTA-----LRKLSQNPSADER 58
              . :: :||:|   : |                          |||  |     :|:| :|  | |
Db       1417 AQKEIEELLDEMQD--------------------NVASIAEHEDTIRRLEENVGAKES 1454

Qy         59 DALQEACLN--KWKIL--------------------SDSLYEQFSKTTRDIELISWFVAA 96
              | |   |   |:|                      ||||     :   |       :|
Db       1455 LLLTEREQNASTLKLLAEAHLEIDELIRKLEDSDRKSDSLQSTIKRLEED------GIAK 1508

Qy         97 QFLLDTTLESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDKG--------------- 141
              : || |  ::   :   | :  ||: : |  :          ||||
Db       1509 EALLLTEKQAHEATRMTLTEALEKNEELLKKI-------HDDDKHILELQFTIQRLEENT 1561

Qy        142 --------KEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLVGEVTFFDF-------- 185
                     :|||| ||  ||  |:          :|          | |
Db       1562 AAKENLLLRERERQNDATTKA--QIESQERNEQLL------------KRFVDVDRKIDLLQ 1607

Qy        186 QSAERKGEISQLKSMLTTTVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQST 245
              : || || | :|  |  :  ||: ||:      |  |   :|   |:      :  : |
Db       1608 DTIERIGENSTIKDALLLSERQEKDAIK------KELVEAGERNEELI------MKIEDT 1655

Qy        246 NFGFAKSLLTRVENALVHLSG------IKL-APKAEAKTVEQEVAES---------SVSE 289
              :       :  ::||:: | |      | | | :  |  |: : :||:          :|:
Db       1656 D-----KKIEHLQNAIIKLEGDIEAKDISLEAAREENDTIRKSLAEAQEKNEELLRKISD 1710

Qy        290 GELPSHM--DTKH------IERIPM---------ASEQAQTVSQHLHAGNLSELGNLNNM 332
              |   |:  ||       | |:           |:::| | ::  :   |    :|
Db       1711 NEYRIHLLQDTAQKLQVDAISRLSSFVMEKQESDAAKRALTEARERNEDLLKRNEDLLKR 1770

Qy        333 NRDLAFHL----------------------LREVSDYFRQSEPHSPISFLLEKAIRWGY 369
              | ||   :                    |   :  ||     :| |      | |
Db       1771 NDDLIKKIEESSKTITQLQETLQRLEGKSTNLEAENQVLRQQATATPPSTAKSSASR--- 1827

Qy        370 LSLPELLREMMSEQNGDALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSD 429
                 :: |   | :||  |:      | :              |: |      | :  ||: :
Db       1828 ---SKITRIHRSPENGHILNGDTRQAEIK-----------PSTG-----TSETIPSIGN 1867

Qy        430 PRSV--EEHVSQ---TSPVDTQSKQDQKPQ 454
              |  :  |:|| |       |  |  | |:||
Db       1868 PPDLNNEKHVEQGEKLQKVLNQKYQSQQPQ 1897
```

RESULT 5
US-10-072-012-838
; Sequence 838, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Esha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.

```
;  APPLICANT:  Wolenc, Adam R.
;  APPLICANT:  Pena, Carol E. A
;  APPLICANT:  Furtak, Katarzyna
;  APPLICANT:  Grosse, William M.
;  APPLICANT:  Alsobrook II, John P.
;  APPLICANT:  Lepley, Denise M.
;  APPLICANT:  Rieger, Daniel K.
;  APPLICANT:  Burgess, Catherine E.
;  TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-258
;  CURRENT APPLICATION NUMBER: US/10/072,012
;  CURRENT FILING DATE:  2002-01-31
;  PRIOR APPLICATION NUMBER: 60/265,102
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: 60/265,514
;  PRIOR FILING DATE: 2001-01-31
;  PRIOR APPLICATION NUMBER: 60/265,517
;  PRIOR FILING DATE: 2001-01-31
;  PRIOR APPLICATION NUMBER: 60/265,412
;  PRIOR FILING DATE: 2001-01-31
;  PRIOR APPLICATION NUMBER: 60/265,395
;  PRIOR FILING DATE: 2001-01-31
;  PRIOR APPLICATION NUMBER: 60/266,406
;  PRIOR FILING DATE: 2001-02-02
;  PRIOR APPLICATION NUMBER: 60/266,767
;  PRIOR FILING DATE: 2001-02-05
;  PRIOR APPLICATION NUMBER: 60/267,057
;  PRIOR FILING DATE: 2001-02-07
;  PRIOR APPLICATION NUMBER: 60/266,975
;  PRIOR FILING DATE: 2001-02-07
;  PRIOR APPLICATION NUMBER: 60/267,459
;  PRIOR FILING DATE: 2001-02-08
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1391
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 838
;   LENGTH: 860
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Myosin tail
US-10-072-012-838

  Query Match             5.5%;  Score 129.5;  DB 12;  Length 860;
  Best Local Similarity   18.5%;  Pred. No. 0.032;
  Matches   98;  Conservative   87;  Mismatches  208;  Indels  137;  Gaps   17;

Qy          7 QIEQLSKPLSD------------------DSICGVYLKLEKSAFRPLRNEFNVAQTALR 47
              ||||| |  :                   |||    |  || | : | :: :  |  |
Db        139 QIEQLQKQKAKAEKEKSQLQAEVDDLLAQLDSITKAKLNAEKKA-KQLESQLSELQVKLD 197

Qy         48 KLSQNPSADERDALQEACLNKWKILSDSLYEQFSKTTRDIE----LISWFVAAQFLLDTT 103
              :| :              ||         |  : |  || :|     :|      :  |::
Db        198 ELQRQ------------LNDLTSQKSRLQSENSDLTRQLEEAEAQVSNLSKLKSQLESQ 244

Qy        104 LESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEE 163
              || |  |||      |:  :  |    :  |: | |  :|: : ::: ||  :     :
Db        245 LEEAKRSLE------EESRERANLQAQLRQLEHDLDSLREQLEEESEAKAELE-----RQ 293

Qy        164 SSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSMLTTTVAQ---------------- 207
              |   | : |       : |: : | |  |: :||  |  :::
Db        294 LSKANAEIQQWR-----SKFESEGALRAEELEELKKKLNQKISELEEAAEAANAKCDSLE 348

Qy        208 --------ERFAIQFKMENAKRCVTQL-----------------DRLSALVSTKCHSLG 241
                     |  :| ::| |    ::|                 | | | : |
Db        349 KTKSRLQSELEDLQIELERANAAASELEKKQKNFDKILAEWKRKVDELQAELDTAQREAR 408

Qy        242 SQSTNFGFAKSLLTRVENALVHLSGIKLAPKAEAKTVEQEVAE--SSVSEGELPSHMDTK 299
              : ||       | |::| |  |      | : | | :: |: :    : ||    |   |
Db        409 NLSTE-------LFRLKNELEELKDQVEALRRENKNLQDEIHDLTDQLGEGGRNVHELEK 461

Qy        300 HIERIPMASEQAQTVSQHLHAG-NLSELG------NLNNMNRDLAFHLLREVSDYFRQSE 352
                |:    :: |   :   |   | |         |: :  ::   |  :  ::    :

Db        462 ARRRLEAEKDELQAALEEAEAALELEESKVLRAQVELSQIRSEIERRLAEKEEEFENTRK 521

Qy        353 PHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDALSTIFNAAGLNHLDQVLLPEVSTPT 412
              |      ::||       |   : :|  | | :        : | : |  :::
Db        522 NH-------QRAIE-------SLQATLEAETKGKAEA--------SRLKKKLEGDINELE 559

Qy        413 VGIESPQTPQAKPSVSDPRSVEEHVSQTSPVDTQSKQDQKPQSSATSALS 462
              : ::       | :    ::|:::  |    : || :::|: :  |    |:
Db        560 IALDHAN----KANAEAQKNVKKYQQQVKELQTQVEEEQRARAEDAREQLA 605


RESULT 6
US-10-037-417-59
; Sequence 59, Application US/10037417
; Publication No. US20040052806A1
; GENERAL INFORMATION:
;  APPLICANT: Kekuda, Ramesh
;  APPLICANT:  Alsobrook II, John P
;  APPLICANT:  Tchernev, Velizar T
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Spytek, Kimberly A
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Grosse, William M
;  APPLICANT:  Lepley, Denise M
;  APPLICANT:  Burgess, Catherine E
;  APPLICANT:  Vernet, Corine A.M.
;  APPLICANT:  Li, Li
;  APPLICANT:  Gorman, Linda
;  APPLICANT:  Edinger, Shlomit R
;  APPLICANT:  Sciore, Paul
;  APPLICANT:  Ellerman, Karen
;  APPLICANT:  Malyankar, Uriel M
;  APPLICANT:  Rothenberg, Mark
;  APPLICANT:  Stone, David J
;  APPLICANT:  Boldog, Ferenc L
;  APPLICANT:  Guo, Xiaojia
;  APPLICANT:  Shenoy, Suresh G
;  APPLICANT:  Anderson, David W
;  APPLICANT:  Padigaru, Muralidhara
;  APPLICANT:  Taupier Jr, Raymond J
;  APPLICANT:  Miller, Charles E
;  APPLICANT:  Eisen, Andrew J
;  TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-235
;  CURRENT APPLICATION NUMBER: US/10/037,417
;  CURRENT FILING DATE:  2002-09-20
;  PRIOR APPLICATION NUMBER: 60/260,018
;  PRIOR FILING DATE: 2001-01-05
;  PRIOR APPLICATION NUMBER: 60/260,360
;  PRIOR FILING DATE: 2001-01-08
;  PRIOR APPLICATION NUMBER: 60/272,411
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: 60/272,817
;  PRIOR FILING DATE: 2001-03-02
;  PRIOR APPLICATION NUMBER: 60/291,186
;  PRIOR FILING DATE: 2001-05-15
;  PRIOR APPLICATION NUMBER: 60/303,231
;  PRIOR FILING DATE: 2001-07-05
;  PRIOR APPLICATION NUMBER: 60/305,060
;  PRIOR FILING DATE: 2001-07-12
;  PRIOR APPLICATION NUMBER: 60/318,405
;  PRIOR FILING DATE: 2001-09-10
;  PRIOR APPLICATION NUMBER: 60/318,700
;  PRIOR FILING DATE: 2001-09-12
;  NUMBER OF SEQ ID NOS: 227
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 59
;   LENGTH: 860
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Mysosin Tail
;   OTHER INFORMATION: Consensus Sequence
```

US-10-037-417-59

Query Match 5.5%; Score 129.5; DB 12; Length 860;
Best Local Similarity 18.5%; Pred. No. 0.032;
Matches 98; Conservative 87; Mismatches 208; Indels 137; Gaps 17;

```
Qy          7 QIEQLSKPLSD------------------DSICGVYLKLEKSAFRPLRNEFNVAQTALR 47
              ||||| |  :                   |||     |  || | : | :: :  |  |
Db        139 QIEQLQKQKAKAEKEKSQLQAEVDDLLAQLDSITKAKLNAEKKA-KQLESQLSELQVKLD 197

Qy         48 KLSQNPSADERDALQEACLNKWKILSDSLYEQFSKTTRDIE----LISWFVAAQFLLDTT 103
              :| :                ||         |  : |  || :|     :|     :  |::
Db        198 ELQRQ-------------LNDLTSQKSRLQSENSDLTRQLEEAEAQVSNLSKLKSQLESQ 244

Qy        104 LESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEE 163
              || |  |||       |:  :  |    :  |: | |  :|: : ::: ||  :     :
Db        245 LEEAKRSLE------EESRERANLQAQLRQLEHDLDSLREQLEEESEAKAELE-----RQ 293

Qy        164 SSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSMLTTTVAQ---------------- 207
              |   | : |       : |: : | |  |: :||  |   :::
Db        294 LSKANAEIQQWR-----SKFESEGALRAEELEELKKKLNQKISELEEAAEAANAKCDSLE 348

Qy        208 --------ERFAIQFKMENAKRCVTQL-----------------DRLSALVSTKCHSLG 241
                     |   :| ::| |    ::|                 | | | : |
Db        349 KTKSRLQSELEDLQIELERANAAASELEKKQKNFDKILAEWKRKVDELQAELDTAQREAR 408

Qy        242 SQSTNFGFAKSLLTRVENALVHLSGIKLAPKAEAKTVEQEVAE--SSVSEGELPSHMDTK 299
              : ||         | |::| |  |       | : | | :: |: :    : ||    |   |
Db        409 NLSTE-------LFRLKNELEELKDQVEALRRENKNLQDEIHDLTDQLGEGGRNVHELEK 461

Qy        300 HIERIPMASEQAQTVSQHLHAG-NLSELG------NLNNMNRDLAFHLLREVSDYFRQSE 352
                |:    ::  |   :   |   | |          |: :  ::   |  :  ::    :
Db        462 ARRRLEAEKDELQAALEEAEAALELEESKVLRAQVELSQIRSEIERRLAEKEEEFENTRK 521

Qy        353 PHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDALSTIFNAAGLNHLDQVLLPEVSTPT 412
              |       ::||       |  : :|  | |  :         : | : |  :::
Db        522 NH-------QRAIE-------SLQATLEAETKGKAEA--------SRLKKKLEGDINELE 559

Qy        413 VGIESPQTPQAKPSVSDPRSVEEHVSQTSPVDTQSKQDQKPQSSATSALS 462
              : ::       | :    ::|:::   |   : || :::|: :  |   |:
Db        560 IALDHAN----KANAEAQKNVKKYQQQVKELQTQVEEEQRAREDAREQLA 605
```

RESULT 7
US-10-080-334-166
; Sequence 166, Application US/10080334
; Publication No. US20040002584A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shimkets, Richard A
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Patturajan, Meera
; APPLICANT: Gangolli, Esha A
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Liu, Xiaohong
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Spaderna, Steven K
; APPLICANT: Zerhusen, Bryan D
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; TITLE OF INVENTION: Using the Same
; FILE REFERENCE: 21402-275
; CURRENT APPLICATION NUMBER: US/10/080,334
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/270,523
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/322,712
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/311,980
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/330,307
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/278,796
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: 60/281,521
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/276,677
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/311,595
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/270,220
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/274,295
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/318,526
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/286,548
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/291,765
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/270,797
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/276,400
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/270,810
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 166
; LENGTH: 860
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-080-334-166

Query Match 5.5%; Score 129.5; DB 15; Length 860;
Best Local Similarity 18.5%; Pred. No. 0.032;
Matches 98; Conservative 87; Mismatches 208; Indels 137; Gaps 17;

```
Qy          7 QIEQLSKPLSD------------------DSICGVYLKLEKSAFRPLRNEFNVAQTALR 47
              ||||| |  :                   |||     |  || | : | :: :  |  |
Db        139 QIEQLQKQKAKAEKEKSQLQAEVDDLLAQLDSITKAKLNAEKKA-KQLESQLSELQVKLD 197

Qy         48 KLSQNPSADERDALQEACLNKWKILSDSLYEQFSKTTRDIE----LISWFVAAQFLLDTT 103
              :| :                ||         |  : |  || :|     :|     :  |::
Db        198 ELQRQ-------------LNDLTSQKSRLQSENSDLTRQLEEAEAQVSNLSKLKSQLESQ 244

Qy        104 LESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEE 163
              || |  |||       |:  :  |    :  |: | |  :|: : ::: ||  :     :
Db        245 LEEAKRSLE------EESRERANLQAQLRQLEHDLDSLREQLEEESEAKAELE-----RQ 293

Qy        164 SSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSMLTTTVAQ---------------- 207
              |   | : |       : |: : | |  |: :||  |   :::
Db        294 LSKANAEIQQWR-----SKFESEGALRAEELEELKKKLNQKISELEEAAEAANAKCDSLE 348

Qy        208 --------ERFAIQFKMENAKRCVTQL-----------------DRLSALVSTKCHSLG 241
                     |   :| ::| |    ::|                 | | | : |
Db        349 KTKSRLQSELEDLQIELERANAAASELEKKQKNFDKILAEWKRKVDELQAELDTAQREAR 408

Qy        242 SQSTNFGFAKSLLTRVENALVHLSGIKLAPKAEAKTVEQEVAE--SSVSEGELPSHMDTK 299
              : ||         | |::| |  |       | : | | :: |: :    : ||    |   |
Db        409 NLSTE-------LFRLKNELEELKDQVEALRRENKNLQDEIHDLTDQLGEGGRNVHELEK 461

Qy        300 HIERIPMASEQAQTVSQHLHAG-NLSELG------NLNNMNRDLAFHLLREVSDYFRQSE 352
                |:    ::  |   :   |   | |          |: :  ::   |  :  ::    :
Db        462 ARRRLEAEKDELQAALEEAEAALELEESKVLRAQVELSQIRSEIERRLAEKEEEFENTRK 521
```

```
Qy        353 PHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDALSTIFNAAGLNHLDQVLLPEVSTPT 412
Db        522 NH-------QRAIE-------SLQATLEAETKGKAEA--------SRLKKKLEGDINELE 559

Qy        413 VGIESPQTPQAKPSVSDPRSVEEHVSQTSPVDTQSKQDQKPQSSATSALS 462
Db        560 IALDHAN----KANAEAQKNVKKYQQQVKELQTQVEEEQRAREDAREQLA 605


RESULT 8
US-10-282-122A-61292
; Sequence 61292, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61292
;   LENGTH: 1294
;   TYPE: PRT
;   ORGANISM: Legionella pneumophila
US-10-282-122A-61292

  Query Match             5.4%;  Score 127;  DB 12;  Length 1294;
  Best Local Similarity   23.1%;  Pred. No. 0.1;
  Matches  104;  Conservative   66;  Mismatches  192;  Indels   88;  Gaps   20;

Qy          7 QIEQLSKPLSDDSICGVYLK----------LEKSAFRPLRNEFNVAQTALRKLS--QNP 53
Db        781 QVQQLEEQLAEKEKENLALKKADKQSQHEKSLDKSAIESLTSELNQLKLELQKQETLQLQ 840

Qy         54 SADERDALQEACLNKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEW 113
Db        841 LKSLRKQIQEQTL-VVEGLKEELQKQKKSNTHQEETIERITKEKSLADSALESLRKEMYE 899

Qy        114 LADLSE-------KHWDHLNPVLPVETL------KSDDDKGKEREQAD-AKVKAFFQLVG 159
Db        900 LTRKNEENQLKLTKQVHSLSEQLEEKQLQIREFEKQLQEKEKRVEQSEKGKASAKRTVAS 959

Qy        160 DSEESSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSMLTTTVAQERFAIQFKMENA 219
Db        960 LREQVSNL---KLQLQQLGEVI-----QEKEKG------SSLISQQSKQIIALQEIIEDQ 1005

Qy        220 KRCVTQLD-RLSALVSTKCHSLGSQST--------NFGFAKSLLTRVENALVHLSGIKLA 270
Db       1006 KRQLEELKIKIQELVSAN-QELGKQNQSLSKENLHNKNTVEDLKKKLNELNVQLEQLHQS 1064

Qy        271 PKAEAKTV-----EQEVAESSVSEGELPSHMDTKH----IERIPMASEQAQTVSQHLHAG 321
Db       1065 SNEKEQTIRKLREELIKKDSSLKQNE-EMQLAQKHLQEEIDRL-----QKEIKQQQLNTN 1118

Qy        322 NL-SELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPEL----- 375
Db       1119 QLESIIAQSKEAEKRYQQALQQKKGIYFARMERVSPI----------YLQIQQIEQKAK 1167

Qy        376 -LREMMSEQNGDALSTIFNAAGL---NHLD 401
Db       1168 ELEERRETEASTAAKTLATKLRLEIKNYLD 1197


RESULT 9
US-10-171-311-48
; Sequence 48, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
;  APPLICANT: Schlegel, Robert
;  APPLICANT:  Chen, Yan
;  APPLICANT:  Zhao, Xumei
;  APPLICANT:  Monahan, John
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Glatt, Karen
;  APPLICANT:  Gannavarapu, Manjula
;  APPLICANT:  Hoersh, Sebastian
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
;  TITLE OF INVENTION:  OF CERVICAL CANCER
;  FILE REFERENCE: MRI-035
;  CURRENT APPLICATION NUMBER: US/10/171,311
;  CURRENT FILING DATE:  2002-06-12
;  PRIOR APPLICATION NUMBER: US 60/298,159
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,155
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/335,936
;  PRIOR FILING DATE: 2001-11-14
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48
;   LENGTH: 752
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-171-311-48

  Query Match             5.4%;  Score 126.5;  DB 14;  Length 752;
  Best Local Similarity   21.9%;  Pred. No. 0.049;
  Matches  111;  Conservative   77;  Mismatches  197;  Indels  123;  Gaps   24;

Qy          5 KHQIEQLSKPLS-----DDSI----------CGVYLKLEKSAFRPLRNEFNVAQTALRK 48
Db        160 KLEVEKLSKRIMALEKLEDAFNKSKQECYSLKCN--LEKERMTTKQLSQELESLKVRIKE 217

Qy         49 LSQNPSADERD--ALQEACLNKWKILS-------DSLYEQFSKTTRDIELISWFVAAQFL 99
Db        218 LEAIESRLEKTEFTLKED-LTKLKTLTVMFVDERKTMSEKLKKTEDKLQAASSQLQVEQN 276

Qy        100 LDTT-----LESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAF 154
Db        277 KVTTVTEKLIEETKRALKSKTDVEEKMY---------------SVTKERDDLKNKLKA- 319
```

```
Qy        155 FQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAER---KGEISQLKSMLTTTVAQERFA 211
Db        320 -----EEEKGNDLLSRVNML----KNRLQSLEAIEKDFLKNKLNQDSGKSTTALHQENNK 370

Qy        212 IQFKMENAKRCVTQLDRLSAL------VSTKCHSLGSQSTN-FGFAKSLLTRVENALVHL 264
Db        371 IKELSQEVERLKLKLKDMKAIEDDLMKTEDEYETLERRYANERDKAQFLSKELEHVKMEL 430

Qy        265 SGIKLAPKAEA-------KTVEQEVAESSVSEGELPSHMDTKHIERIP--MASEQAQTVS 315
Db        431 AKYKLAEKTETSHEQWLFKRLQEEEAKS----GHLSREVDALK-EKIHEYMATEDLICHL 485

Qy        316 QHLHAGNLSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPEL 375
Db        486 QGDHSVCKKKLNQQENRNRDLG----REIENLTKELERYRHFS----KSLR------PSL 531

Qy        376 LREMMSEQNGDALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPRSVEE 435
Db        532 --------NGRRISD----------PQVFSKEVQTEAVDNEPPDYKSLIP-------LER 566

Qy        436 HVSQTSPVDTQSKQDQKPQSSATSALSW 463
Db        567 AVINGQLYEESENQDEDPNDEG-SVLSF 593
```

RESULT 10
US-10-309-851-24
; Sequence 24, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98,723-F-US
; CURRENT APPLICATION NUMBER: US/10/309,851
; CURRENT FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 785
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-851-24

Query Match 5.4%; Score 126.5; DB 14; Length 785;
Best Local Similarity 21.9%; Pred. No. 0.052;
Matches 111; Conservative 77; Mismatches 197; Indels 123; Gaps 24;

```
Qy          5 KHQIEQLSKPLS-----DDSI-----------CGVYLKLEKSAFRPLRNEFNVAQTALRK 48
Db         64 KLEVEKLSKRIMALEKLEDAFNKSKQECYSLKCN--LEKERMTTKQLSQELESLKVRIKE 121

Qy         49 LSQNPSADERD--ALQEACLNKWKILS-------DSLYEQFSKTTRDIELISWFVAAQFL 99
Db        122 LEAIESRLEKTEFTLKED-LTKLKTLTVMFVDERKTMSEKLKKTEDKLQAASSQLQVEQN 180

Qy        100 LDTT-----LESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAF 154
Db        181 KVTTVTEKLIEETKRALKSKTDVEEKMY---------------SVTKERDDLKNKLKA- 223

Qy        155 FQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAER---KGEISQLKSMLTTTVAQERFA 211
Db        224 -----EEEKGNDLLSRVNML----KNRLQSLEAIEKDFLKNKLNQDSGKSTTALHQENNK 274

Qy        212 IQFKMENAKRCVTQLDRLSAL------VSTKCHSLGSQSTN-FGFAKSLLTRVENALVHL 264
Db        275 IKELSQEVERLKLKLKDMKAIEDDLMKTEDEYETLERRYANERDKAQFLSKELEHVKMEL 334

Qy        265 SGIKLAPKAEA-------KTVEQEVAESSVSEGELPSHMDTKHIERIP--MASEQAQTVS 315
Db        335 AKYKLAEKTETSHEQWLFKRLQEEEAKS----GHLSREVDALK-EKIHEYMATEDLICHL 389

Qy        316 QHLHAGNLSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPEL 375
Db        390 QGDHSVLQKKLNQQENRNRDLG----REIENLTKELERYRHFS----KSLR------PSL 435

Qy        376 LREMMSEQNGDALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPRSVEE 435
Db        436 --------NGRRISD----------PQVFSKEVQTEAVDNEPPDYKSLIP-------LER 470

Qy        436 HVSQTSPVDTQSKQDQKPQSSATSALSW 463
Db        471 AVINGQLYEESENQDEDPNDEG-SVLSF 497
```

RESULT 11
US-10-108-260A-3569
; Sequence 3569, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3569
; LENGTH: 824
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3569

Query Match 5.4%; Score 126.5; DB 15; Length 824;
Best Local Similarity 21.9%; Pred. No. 0.056;
Matches 111; Conservative 77; Mismatches 197; Indels 123; Gaps 24;

```
Qy          5 KHQIEQLSKPLS-----DDSI-----------CGVYLKLEKSAFRPLRNEFNVAQTALRK 48
Db        105 KLEVEKLSKRIMALEKLEDAFNKSKQECYSLKCN--LEKERMTTKQLSQELESLKVRIKE 162

Qy         49 LSQNPSADERD--ALQEACLNKWKILS-------DSLYEQFSKTTRDIELISWFVAAQFL 99
Db        163 LEAIESRLEKTEFTLKED-LTKLKTLTVMFVDERKTMSEKLKKTEDKLQAASSQLQVEQN 221

Qy        100 LDTT-----LESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAF 154
Db        222 KVTTVTEKLIEETKRALKSKTDVEEKMY---------------SVTKERDDLKNKLKA- 264

Qy        155 FQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAER---KGEISQLKSMLTTTVAQERFA 211
Db        265 -----EEEKGNDLLSRVNML----KNRLQSLEAIEKDFLKNKLNQDSGKSTTALHQENNK 315

Qy        212 IQFKMENAKRCVTQLDRLSAL------VSTKCHSLGSQSTN-FGFAKSLLTRVENALVHL 264
Db        316 IKELSQEVERLKLKLKDMKAIEDDLMKTEDEYETLERRYANERDKAQFLSKELEHVKMEL 375

Qy        265 SGIKLAPKAEA-------KTVEQEVAESSVSEGELPSHMDTKHIERIP--MASEQAQTVS 315
Db        376 AKYKLAEKTETSHEQWLFKRLQEEEAKS----GHLSREVDALK-EKIHEYMATEDLICHL 430

Qy        316 QHLHAGNLSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPEL 375
Db        431 QGDHSVLQKKLNQQENRNRDLG----REIENLTKELERYRHFS----KSLR------PSL 476

Qy        376 LREMMSEQNGDALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPRSVEE 435
Db        477 --------NGRRISD----------PQVFSKEVQTEAVDNEPPDYKSLIP-------LER 511

Qy        436 HVSQTSPVDTQSKQDQKPQSSATSALSW 463
Db        512 AVINGQLYEESENQDEDPNDEG-SVLSF 538
```

RESULT 12

```
US-10-309-851-14
; Sequence 14, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
;  APPLICANT: Saus, Juan
;  APPLICANT:  Revert-Ros, Francisco
;  TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
;  TITLE OF INVENTION:  Interact with Goodpasture Antigen Binding Protein
;  FILE REFERENCE: 98,723-F-US
;  CURRENT APPLICATION NUMBER: US/10/309,851
;  CURRENT FILING DATE:  2002-12-04
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
;   LENGTH: 1133
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-309-851-14

  Query Match             5.4%;  Score 126.5;  DB 14;  Length 1133;
  Best Local Similarity   21.9%;  Pred. No. 0.091;
  Matches  111;  Conservative   77;  Mismatches  197;  Indels  123;  Gaps   24;

Qy          5 KHQIEQLSKPLS-----DDSI-----------CGVYLKLEKSAFRPLRNEFNVAQTALRK 48
                | ::|:|||  :          :|:             |    |: |:    :  |    |     :  :::
Db        414 KLEVEKLSKRIMALEKLEDAFNKSKQECYSLKCN--LEKERMTTKQLSQELESLKVRIKE 471

Qy         49 LSQNPSADERD--ALQEACLNKWKILS-------DSLYEQFSKTTRDIELISWFVAAQFL 99
                    |    |  |:     |:|  | | | |:          :: |:  ||    :: |  :  :
Db        472 LEAIESRLEKTEFTLKED-LTKLKTLTVMFVDERKTMSEKLKKTEDKLQAASSQLQVEQN 530

Qy        100 LDTT-----LESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAF 154
                  ||        :|     :|:    |: || :                         |||:      |:||
Db        531 KVTTVTEKLIEETKRALKSKTDVEEKMY---------------SVTKERDDLKNKLKA- 573

Qy        155 FQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAER---KGEISQLKSMLTTTVAQERFA 211
                      : |: :  | :  |    |      :         :: |:     | :::|        || : ||
Db        574 -----EEEKGNDLLSRVNML----KNRLQSLEAIEKDFLKNKLNQDSGKSTTALHQENNK 624

Qy        212 IQFKMENAKRCVTQLDRLSAL------VSTKCHSLGSQSTN-FGFAKSLLTRVENALVHL 264
                |:   :   :|    :|  : |:          :   :|   :    |      |: |     :|: :  |
Db        625 IKELSQEVERLKLKLKDMKAIEDDLMKTEDEYETLERRYANERDKAQFLSKELEHVKMEL 684

Qy        265 SGIKLAPKAEA-------KTVEQEVAESSVSEGELPSHMDTKHIERIP--MASEQAQTVS 315
                :  ||| |  |            |  :::|  |:|     |  |     :|      |:|      ||:|
Db        685 AKYKLAEKTETSHEQWLFKRLQEEEAKS----GHLSREVDALK-EKIHEYMATEDLICHL 739

Qy        316 QHLHAGNLSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPEL 375
                |   |:      :|     | |||      ||: :   :: |  :    |      |::|           | |
Db        740 QGDHSVLQKKLNQQENRNRDLG----REIENLTKELERYRHFS----KSLR------PSL 785

Qy        376 LREMMSEQNGDALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPRSVEE 435
                            ||  :|              ||   ||  |   |   |    |              |          :|
Db        786 --------NGRRISD----------PQVFSKEVQTEAVDNEPPDYKSLIP-------LER 820

Qy        436 HVSQTSPVDTQSKQDQKPQSSATSALSW 463
                  |           :         ||: |        | ||:
Db        821 AVINGQLYEESENQDEDPNDEG-SVLSF 847

RESULT 13
US-10-309-851-16
; Sequence 16, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
;  APPLICANT: Saus, Juan
;  APPLICANT:  Revert-Ros, Francisco
;  TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
;  TITLE OF INVENTION:  Interact with Goodpasture Antigen Binding Protein
;  FILE REFERENCE: 98,723-F-US
;  CURRENT APPLICATION NUMBER: US/10/309,851
;  CURRENT FILING DATE:  2002-12-04
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
;   LENGTH: 1133
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-309-851-16

  Query Match             5.4%;  Score 126.5;  DB 14;  Length 1133;
  Best Local Similarity   21.9%;  Pred. No. 0.091;
  Matches  111;  Conservative   77;  Mismatches  197;  Indels  123;  Gaps   24;

Qy          5 KHQIEQLSKPLS-----DDSI-----------CGVYLKLEKSAFRPLRNEFNVAQTALRK 48
                | ::|:|||  :          :|:             |    |: |:    :  |    |     :  :::
Db        414 KLEVEKLSKRIMALEKLEDAFNKSKQECYSLKCN--LEKERMTTKQLSQELESLKVRIKE 471

Qy         49 LSQNPSADERD--ALQEACLNKWKILS-------DSLYEQFSKTTRDIELISWFVAAQFL 99
                    |    |  |:     |:|  | | | |:          :: |:  ||    :: |  :  :
Db        472 LEAIESRLEKTEFTLKED-LTKLKTLTVMFVDERKTMSEKLKKTEDKLQAASSQLQVEQN 530

Qy        100 LDTT-----LESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAF 154
                  ||        :|     :|:    |: || :                         |||:      |:||
Db        531 KVTTVTEKLIEETKRALKSKTDVEEKMY---------------SVTKERDDLKNKLKA- 573

Qy        155 FQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAER---KGEISQLKSMLTTTVAQERFA 211
                      : |: :  | :  |    |      :         :: |:     | :::|        || : ||
Db        574 -----EEEKGNDLLSRVNML----KNRLQSLEAIEKDFLKNKLNQDSGKSTTALHQENNK 624

Qy        212 IQFKMENAKRCVTQLDRLSAL------VSTKCHSLGSQSTN-FGFAKSLLTRVENALVHL 264
                |:   :   :|    :|  : |:          :   :|   :    |      |: |     :|: :  |
Db        625 IKELSQEVERLKLKLKDMKAIEDDLMKTEDEYETLERRYANERDKAQFLSKELEHVKMEL 684

Qy        265 SGIKLAPKAEA-------KTVEQEVAESSVSEGELPSHMDTKHIERIP--MASEQAQTVS 315
                :  ||| |  |            |  :::|  |:|     |  |     :|      |:|      ||:|
Db        685 AKYKLAEKTETSHEQWLFKRLQEEEAKS----GHLSREVDALK-EKIHEYMATEDLICHL 739

Qy        316 QHLHAGNLSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPEL 375
                |   |:      :|     | |||      ||: :   :: |  :    |      |::|           | |
Db        740 QGDHSVLQKKLNQQENRNRDLG----REIENLTKELERYRHFS----KSLR------PSL 785

Qy        376 LREMMSEQNGDALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPRSVEE 435
                            ||  :|              ||   ||  |   |   |    |              |          :|
Db        786 --------NGRRISD----------PQVFSKEVQTEAVDNEPPDYKSLIP-------LER 820

Qy        436 HVSQTSPVDTQSKQDQKPQSSATSALSW 463
                  |           :         ||: |        | ||:
Db        821 AVINGQLYEESENQDEDPNDEG-SVLSF 847

RESULT 14
US-10-309-851-12
; Sequence 12, Application US/10309851
; Publication No. US20030108554A1
; GENERAL INFORMATION:
;  APPLICANT: Saus, Juan
;  APPLICANT:  Revert-Ros, Francisco
;  TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
;  TITLE OF INVENTION:  Interact with Goodpasture Antigen Binding Protein
;  FILE REFERENCE: 98,723-F-US
;  CURRENT APPLICATION NUMBER: US/10/309,851
;  CURRENT FILING DATE:  2002-12-04
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
;   LENGTH: 1135
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-309-851-12

  Query Match             5.4%;  Score 126.5;  DB 14;  Length 1135;
  Best Local Similarity   21.9%;  Pred. No. 0.091;
  Matches  111;  Conservative   77;  Mismatches  197;  Indels  123;  Gaps   24;

Qy          5 KHQIEQLSKPLS-----DDSI-----------CGVYLKLEKSAFRPLRNEFNVAQTALRK 48
```

```
Db        414 KLEVEKLSKRIMALEKLEDAFNKSKQECYSLKCN--LEKERMTTKQLSQELESLKVRIKE 471

Qy         49 LSQNPSADERD--ALQEACLNKWKILS-------DSLYEQFSKTTRDIELISWFVAAQFL 99

Db        472 LEAIESRLEKTEFTLKED-LTKLKTLTVMFVDERKTMSEKLKKTEDKLQAASSQLQVEQN 530

Qy        100 LDTT-----LESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAF 154

Db        531 KVTTVTEKLIEETKRALKSKTDVEEKMY---------------SVTKERDDLKNKLKA- 573

Qy        155 FQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAER---KGEISQLKSMLTTTVAQERFA 211

Db        574 -----EEEKGNDLLSRVNML----KNRLQSLEAIEKDFLKNKLNQDSGKSTTALHQENNK 624

Qy        212 IQFKMENAKRCVTQLDRLSAL------VSTKCHSLGSQSTN-FGFAKSLLTRVENALVHL 264

Db        625 IKELSQEVERLKLKLKDMKAIEDDLMKTEDEYETLERRYANERDKAQFLSKELEHVKMEL 684

Qy        265 SGIKLAPKAEA-------KTVEQEVAESSVSEGELPSHMDTKHIERIP--MASEQAQTVS 315

Db        685 AKYKLAEKTETSHEQWLFKRLQEEEAKS----GHLSREVDALK-EKIHEYMATEDLICHL 739

Qy        316 QHLHAGNLSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPEL 375

Db        740 QGDHSVLQKKLNQQENRNRDLG----REIENLTKELERYRHFS----KSLR------PSL 785

Qy        376 LREMMSEQNGDALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPRSVEE 435

Db        786 --------NGRRISD----------PQVFSKEVQTEAVDNEPPDYKSLIP-------LER 820

Qy        436 HVSQTSPVDTQSKQDQKPQSSATSALSW 463

Db        821 AVINGQLYEESENQDEDPNDEG-SVLSF 847


RESULT 15
US-10-157-223-7
; Sequence 7, Application US/10157223
; Publication No. US20030036510A1
;   GENERAL INFORMATION:
;        APPLICANT: Bandman, Olga
;                   Hillman, Jennifer L.
;                   Au-Young, Janice
;                   Zweiger, Gary B.
;        TITLE OF INVENTION: NOVEL HUMAN CELL DIVISION CYCLE PROTEINS
;        NUMBER OF SEQUENCES: 9
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Incyte Pharmaceuticals, Inc.
;             STREET: 3174 Porter Drive
;             CITY: Palo Alto
;             STATE: CA
;             COUNTRY: U.S.
;             ZIP: 94304
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ Version 1.5
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/157,223
;             FILING DATE: 28-May-2002
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/712,708
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Billings, Lucy J.
;             REGISTRATION NUMBER: 36,749
;             REFERENCE/DOCKET NUMBER: PF-0122 US
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 415-855-0555
;             TELEFAX: 415-845-4166
;   INFORMATION FOR SEQ ID NO: 7:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 506 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        IMMEDIATE SOURCE:
;             LIBRARY: GenBank
;             CLONE: 1077057
;        SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-10-157-223-7

  Query Match             5.4%;  Score 126;  DB 14;  Length 506;
  Best Local Similarity   19.1%;  Pred. No. 0.03;
  Matches   90;  Conservative   71;  Mismatches  161;  Indels  148;  Gaps   17;

Qy         15 LSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACLNKWKILSD 74

Db         13 LSDDSDVEVHPNVDKKSF---------------------------------IKWK--QQ 36

Qy         75 SLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNPVLPVETL 134

Db         37 SIHEQRFKRNQDIKNLETQVDMYSHLNKRVDRI------LSNLPESSLTDLPAVTKFLNA 90

Qy        135 KSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAERKGEI 194

Db         91 NFDKMEKSKGENVDPEIATYNEMVEDLFE---------QLA----------KDLDKEGKD 131

Qy        195 SQLKSMLTTTVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLL 254

Db        132 SKSPSLIRDAILKHRAKIDSVTVEAKK------KLDELYKEKNAHISSEDIHTGFDSSFM 185

Qy        255 TRVENALVHLSGI---KLAPKAEAKTVEQEVAESSVSE------------------GEL 292

Db        186 NKQKGGAKPLEATPSEALSSAAES-NILNKLAKSSVPQTFIDFKDDPMKLAKETEEFGKI 244

Qy        293 PSHMDTKH----IERIPMASEQ------AQTVSQHLHAG---------------------- 321

Db        245 SINEYSKSQKFLLEHLPIISEQQKDALMMKAFEYQLHGDDKMTLQVIHQSELMAYIKEIY 304

Qy        322 --------NLSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLP 373

Db        305 DMKKIPYLNPMELSNVINM-------FFEKV--IFNKDKPMGKESFL--RSVQEKFLHIQ 353

Qy        374 E----LLREMMSEQNGDALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQ 419

Db        354 KRSKILQQEEMDESNAEGVETI----QLKSLDDSTELEVNLPDFNSKDPE 399


Search completed: July  3, 2004, 02:52:30
Job time : 83 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 3, 2004, 02:16:54 ; Search time 27 Seconds
(without alignments)
1649.507 Million cell updates/sec

Title: US-10-780-347-2
Perfect score: 2343
Sequence: 1 MPLSKHQIEQLSKPLSDDSI.....DTQSKQDQKPQSSATSALSW 463

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	241.5	10.3	351	2	AC3090	conserved hypothet
2	241.5	10.3	351	2	G98196	hypothetical prote
3	209	8.9	366	2	D83350	hypothetical prote
4	152	6.5	2245	2	T18278	myosin heavy chain
5	143	6.1	884	2	T20405	hypothetical prote
6	141.5	6.0	344	2	G83634	hypothetical prote
7	141.5	6.0	1510	2	T16927	hypothetical prote
8	134.5	5.7	1179	2	E98012	hypothetical prote
9	134	5.7	1051	2	T18302	apsB protein - Eme
10	129.5	5.5	1133	2	T22976	hypothetical prote
11	127.5	5.4	2471	2	T42977	large tegument pro
12	126.5	5.4	1023	2	AE1643	ATP-dependent dsDN
13	126	5.4	506	2	S50914	cell division cont
14	125.5	5.4	1287	2	T22235	hypothetical prote
15	125.5	5.4	1992	2	A47297	myosin heavy chain
16	125	5.3	1194	2	T37503	probable chromosom
17	124.5	5.3	550	2	T40379	hypothetical prote
18	124.5	5.3	1940	1	S04090	myosin heavy chain
19	124.5	5.3	1940	1	A24922	myosin heavy chain
20	124.5	5.3	1964	2	A59282	nonmuscle myosin I
21	123.5	5.3	867	2	C71638	DNA polymerase I (
22	123.5	5.3	1051	2	T18351	lmp1 protein - Myc
23	123	5.2	1935	1	A37102	myosin beta heavy
24	123	5.2	1935	1	*S06006*	*myosin beta heavy*
25	123	5.2	2022	2	T43214	ovt1 protein - nem
26	122.5	5.2	2469	2	H36812	hypothetical prote
27	122	5.2	1934	2	I48153	myosin heavy chain
28	122	5.2	2109	2	E89066	protein H05009.1 [
29	122	5.2	2109	2	T33247	hypothetical prote
30	121.5	5.2	1179	2	G95144	conserved hypothet
31	121.5	5.2	3225	2	I52300	giantin - human
32	121.5	5.2	3259	1	A56539	giantin - human
33	120.5	5.1	999	2	JC5278	oxygen-regulated p
34	120.5	5.1	1938	2	A59293	skeletal myosin he
35	120	5.1	650	2	S44806	F10E9.6 protein -
36	120	5.1	828	2	F96535	hypothetical prote
37	120	5.1	1034	2	T32297	hypothetical prote
38	119.5	5.1	1066	1	A48669	kinesin-related pr
39	119.5	5.1	1690	2	T13030	microtubule bindin
40	119	5.1	1023	2	AE1280	ATP-dependent dsDN
41	119	5.1	1356	2	S32763	kinectin 1 - human
42	119	5.1	1935	2	A59286	myosin heavy chain
43	119	5.1	2104	2	T38774	myosin-3 heavy cha
44	118.5	5.1	1920	2	A53188	pericentrin - mous
45	118.5	5.1	1937	2	I38055	myosin heavy chain

ALIGNMENTS

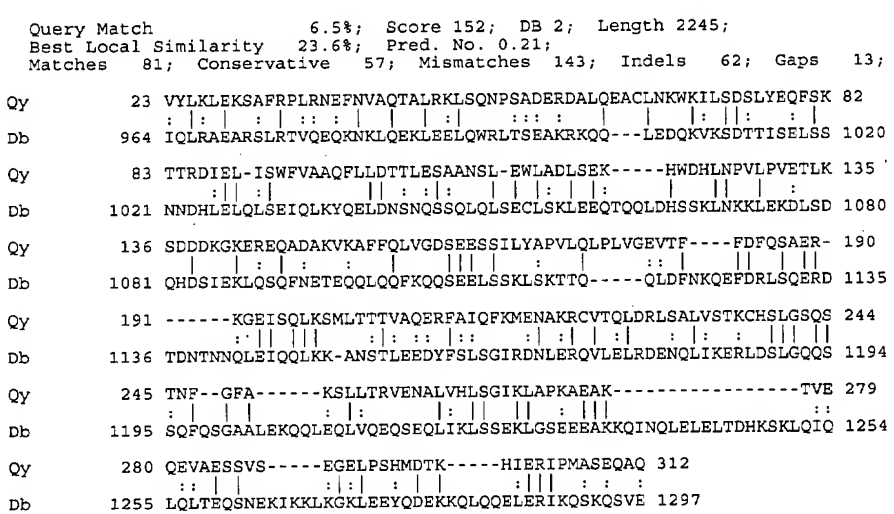
RESULT 1
AC3090
conserved hypothetical protein Atu4343 [imported] - Agrobacterium tumefaciens (strain C
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AC3090
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClel
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AC3090
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-351 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL45137.1; PID:g17742810; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu4343
A;Map position: linear chromosome

Query Match 10.3%; Score 241.5; DB 2; Length 351;
Best Local Similarity 23.6%; Pred. No. 2.3e-08;
Matches 96; Conservative 64; Mismatches 162; Indels 85; Gaps 15;

```
Qy          5 KHQIEQLSKPLSDDSICGVYLKLE---KSAFRPLRNEFNVAQTALRKLSQNPSADERDAL 61
              |  || |          ||  :: :   :  :  :::| | |:           |:|| |
Db          7 KRDIEFLGN-------CGDNIRNDSRTREIYYRIKDERNQAR----------AEERAAS 48

Qy         62 QEACL---NKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLS 118
              :  |   : |  :|:   :     ::|:|:::|   |   |           |   ||
Db         49 PQDNLKISSSWDSVSNLGLQIIYSESKDVEILAWLAEASLRL-RGFHGLREIYELCGDLF 107

Qy        119 EKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLV- 177
              |||  |  :        |||:       :  |   |  |  |  | : |  |:    |:
Db        108 YNHWDSLRSI-------SDDND-------EEKFAPFAGLNGIGSEGT-LVQPLRLASLIP 152

Qy        178 ----GEVTFFDFQSAERKGEISQLKSMLTTTVAQERFAIQFKMENAKRCVTQLDRLSALV 233
                  || : :|||  |:|  |  : : :        |:   :     |::  | ::|::
Db        153 GGKFGEHSLWDFQLAQRPNESKRREELYRIASEAGVAAMSSHLAAVNTCLSSFDAITAVL 212

Qy        234 STKCHSLGSQSTNFGFAKSLLTRVENALVHLSGIKLAPKAEAKTV---EQEVAESSVSEG 290
              | :|      |:|          : | |:  :       |  :|:   :|| |
Db        213 SERCGQAAPPSSN----------IRNTLIEAA-------AAIRTLGGRDQEPAP------ 249

Qy        291 ELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRDLAFHLLREVSDYFRQ 350
                      :|: |  :  | |       |   |         :|| ||  |  |: |||:
Db        250 ----------VEQTPAIA--AGTDESGQSAARTSPASPEGISSRDEAFETLLSVARYFRR 297
```

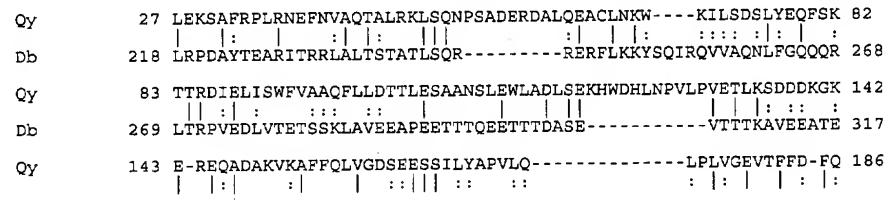
```
Qy        351 SEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDALSTIFNAAGL 397
              :|||||||  :|  :| | :   ||| |:: |    | : :  |||:
Db        298 TEPHSPISLSIETLVRRGRMDFSELLAELLPET--QARNAVLTAAGI 342


RESULT 2
G98196
hypothetical protein AGR_L_1042 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: G98196
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: G98196
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-351 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89097.1; PID:g15158901; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_1042
A;Map position: linear chromosome

  Query Match             10.3%;  Score 241.5;  DB 2;  Length 351;
  Best Local Similarity   23.6%;  Pred. No. 2.3e-08;
  Matches   96;  Conservative   64;  Mismatches  162;  Indels   85;  Gaps   15;

Qy          5 KHQIEQLSKPLSDDSICGVYLKLE---KSAFRPLRNEFNVAQTALRKLSQNPSADERDAL 61
              |  || |          ||  :: :   :  :  :::| | |:           |:|| |
Db          7 KRDIEFLGN-------CGDNIRNDSRTREIYYRIKDERNQAR----------AEERAAS 48

Qy         62 QEACL---NKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLS 118
                :  |   : |  :|:   :     ::|:|:::|   |   |           |   ||
Db         49 PQDNLKISSSWDSVSNLGLQIIYSESKDVEILAWLAEASLRL-RGFHGLREIYELCGDLF 107

Qy        119 EKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLV- 177
              |||  |  :      |||:       : |   |  | |   | : |  |:    |:
Db        108 YNHWDSLRSI-------SDDND-------EEKFAPFAGLNGIGSEGT-LVQPLRLASLIP 152

Qy        178 ----GEVTFFDFQSAERKGEISQLKSMLTTTVAQERFAIQFKMENAKRCVTQLDRLSALV 233
                  || : :||| |:|  |  : : :          |:   :     |::  | ::|::
Db        153 GGKFGEHSLWDFQLAQRPNESKRREELYRIASEAGVAAMSSHLAAVNTCLSSFDAITAVL 212

Qy        234 STKCHSLGSQSTNFGFAKSLLTRVENALVHLSGIKLAPKAEAKTV---EQEVAESSVSEG 290
              | :|      |:|          : | |:  :       |  :|:   :|| |
Db        213 SERCGQAAPPSSN----------IRNTLIEAA-------AAIRTLGGRDQEPAP------ 249

Qy        291 ELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRDLAFHLLREVSDYFRQ 350
                        :|: |  :  | |      |    |          :|| ||  |  |: |||:
Db        250 ----------VEQTPAIA--AGTDESGQSAARTSPASPEGISSRDEAFETLLSVARYFRR 297

Qy        351 SEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDALSTIFNAAGL 397
              :|||||||  :|  :| | :   ||| |:: |    | : :  |||:
Db        298 TEPHSPISLSIETLVRRGRMDFSELLAELLPET--QARNAVLTAAGI 342


RESULT 3
D83350
hypothetical protein PA2360 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: D83350
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: D83350
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-366 <STO>
A;Cross-references: GB:AE004662; GB:AE004091; NID:g9948395; PIDN:AAG05748.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA2360

  Query Match              8.9%;  Score 209;  DB 2;  Length 366;
  Best Local Similarity   21.5%;  Pred. No. 3.1e-06;
  Matches   91;  Conservative   73;  Mismatches  151;  Indels  108;  Gaps   16;

Qy          1 MPLSKH--QIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADER 58
              :||| :   :| | :|:      ||   |: :   |             ||:| :   :
Db          3 LPLSGNALSLEVLLEPIDPGQACGPSLRYDPDYDR-----------LRELRREDDSSLP 50

Qy         59 DALQEACLNK--WKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLAD 116
                : :|   :  |  :       |   : ::|: | :|   | :|     |      :|  ||:
Db         51 TGVWQAEAKRADWAAVEQLASELLQRRSKDLMLAAWLGEA-WLQRGGLGGLQRALVLLAE 109

Qy        117 LSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPL 176
              | |:: : ::|          | |:        |  :: : :|:              :|||
Db        110 LCERYPEEVHP------QAQDGDQSWRVPPIDWLLRRYAELLH-----------TRLPL 151

Qy        177 VG-----EVTFFDFQSAERK----GEISQLKSMLTTTVAQERFAIQFKMENAKR------ 221
              :|      |:| : :|  :|:     |:     |: |      |::      |:: | |
Db        152 MGQGAFAEITLYAWQRLQRQQVASGDSKSAKAALEAAQLQQK-----KLDEALRAEPLVQ 206

Qy        222 ----------CVTQL-------DRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVHL 264
                       |  ||       ||    ::  | |           |: |     |   :|
Db        207 WQRKQASLLACQQQLLRLEQWCDRCLGELAPSCQPLREV-----IAQWLALLKEFIAMH- 260

Qy        265 SGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLS 324
                    |:|      : |||:  |||       ||    | :| ::      :
Db        261 ------PQAPLSEEQPPVAEADASEG------DTDGEESVPASAPSGPAGAP-------- 300

Qy        325 ELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQN 384
                       :|: |:  |  ::||  ::|||||: :|:::|: ||    | ||| |::   |
Db        301 -------TSREDAYRQLLLIADYLARTEPHSPVPYLIKRAVEWGNKPLSELLAELI---N 350

Qy        385 GDA 387
                |:
Db        351 ADS 353


RESULT 4
T18278
myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C;Accession: T18278; T30579
R;Hammer III, J.A.; Jung, G.
J. Biol. Chem. 271, 7120-7127, 1996
A;Title: The sequence of the dictyostelium myo J heavy chain gene predicts a novel, dime
A;Reference number: Z18854; MUID:96215148; PMID:8636147
A;Accession: T18278
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2245 <HAM>
A;Cross-references: EMBL:U42409; NID:g1150765; PID:g1150766; PIDN:AAA85186.1
R;Titus, M.A.; Kuspa, A.; Loomis, W.F.
Proc. Natl. Acad. Sci. U.S.A. 91, 9446-9450, 1994
A;Title: Discovery of myosin genes by physical mapping in Dictyostelium.
A;Reference number: Z20873; MUID:95023928; PMID:7937787
A;Accession: T30579
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-190,'F',192-283,'T',285-290,'R',292-331,'IE','WM',337-338,'LK',342,'YRMKS'
A;Cross-references: EMBL:L35322; NID:g1039360; PID:g1039361; PIDN:AAA79858.1
C;Genetics:
A;Gene: myoJ
A;Introns: 257/3; 307/3
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: ATP; P-loop
F;84-809/Domain: myosin motor domain homology <MMO>
```

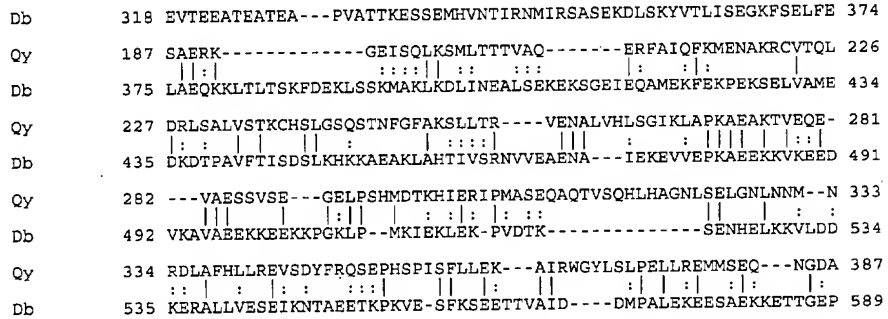
Query Match 6.5%; Score 152; DB 2; Length 2245;
Best Local Similarity 23.6%; Pred. No. 0.21;
Matches 81; Conservative 57; Mismatches 143; Indels 62; Gaps 13;



RESULT 5
T20405
hypothetical protein E01G4.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T20405; T22248
R;Sims, M.
submitted to the EMBL Data Library, December 1996
A;Reference number: Z19269
A;Accession: T20405
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-884 <WIL>
A;Cross-references: EMBL:Z83223; PIDN:CAB05719.1; GSPDB:GN00020; CESP:E01G4.6
A;Experimental source: clone E01G4
R;Sims, M.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19536
A;Accession: T22248
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-884 <WI2>
A;Cross-references: EMBL:Z81538; PIDN:CAB04387.1; GSPDB:GN00020; CESP:E01G4.6
A;Experimental source: clone F45H10
C;Genetics:
A;Gene: CESP:E01G4.6
A;Map position: 2
A;Introns: 24/3; 77/1; 745/3

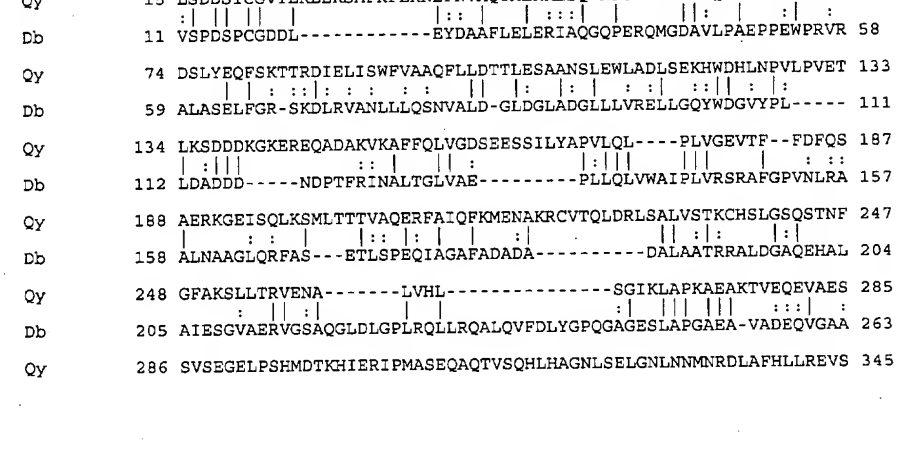
Query Match 6.1%; Score 143; DB 2; Length 884;
Best Local Similarity 20.7%; Pred. No. 0.21;
Matches 103; Conservative 99; Mismatches 165; Indels 130; Gaps 25;





RESULT 6
G83634
hypothetical protein PA0082 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: G83634
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: G83634
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-344 <STO>
A;Cross-references: GB:AE004447; GB:AE004091; NID:g9945902; PIDN:AAG03472.1; GSPDB:GN00
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA0082

Query Match 6.0%; Score 141.5; DB 2; Length 344;
Best Local Similarity 21.1%; Pred. No. 0.068;
Matches 89; Conservative 70; Mismatches 144; Indels 119; Gaps 18;



```
Db        264 PVAAVAAPA----------PRAS-----------GEIA--------NREDVLRQLDRLL 293

Qy        346 DYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDALSTIFNAAGLNHLDQVLL 405

Db        294 EYYVRHEPSSPVPVLLKRAKTLVTADFAEIVRNLIPD------------GISQFETLRG 340

Qy        406 PE 407

Db        341 PE 342
```

RESULT 7
T16927
hypothetical protein T23F2.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T16927
R;Du, Z.
submitted to the EMBL Data Library, October 1995
A;Description: The sequence of C. elegans cosmid T23F2.
A;Reference number: Z18608
A;Accession: T16927
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1510 <DUZ>
A;Cross-references: EMBL:U39649; NID:g1049370; PID:g1049372; PIDN:AAA80384.1; CESP:T23F2
C;Genetics:
A;Gene: CESP:T23F2.2
A;Introns: 20/2; 81/1; 119/2; 211/3; 236/3; 290/2; 344/1; 396/3; 439/3; 465/3; 556/3; 60

  Query Match 6.0%; Score 141.5; DB 2; Length 1510;
  Best Local Similarity 20.0%; Pred. No. 0.56;
  Matches 112; Conservative 89; Mismatches 168; Indels 191; Gaps 26;

```
Qy          9 EQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACLNK 68

Db        619 QQLSEHESD------FLSTQQT-LHDLKKEYSAKNTTL--------VDKFKEVEEILLAK 663

Qy         69 WKILSDSLYEQFSK----TTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDH 124

Db        664 TELV-DALTKQLENIRKDQTRELS----------LKQSERDQYKKSLEEMTFIAEK---- 708

Qy        125 LNPVLPVETLKSDDDKGKEREQADAKVK---AFFQ-----LVGDS-------------- 161

Db        709 -VPILEAEILQLSKDK----NEITARLKHDQEYFEDELAKLLNDSMNIKKERDDYLTEHI 763

Qy        162 --EESSILYAPVLQLPLVGEVTFFDFQSAER-KGEISQLKSMLTTTVAQERFAIQFKMEN 218

Db        764 RANESMI---------------------ERLKLEISGLKKDLENQKMQAHLQ---KSEL 798

Qy        219 AKRCVTQLDRLSALVS--------TKCHSLGSQSTNFGFAK------------------ 251

Db        799 EKKLLSSIDHVSQLQSRVNHSQRDVECQAIPRQINKYVGCKPNVKNKETIIEKGALFDEN 858

Qy        252 ---------------SLLTRVENALVHL----------SGIKLAPKAEAKTVEQEVAESS 286

Db        859 EERLRICKAELETTRRQVTVLQQKLVSIIQQQSSQKIKKRIAVVEDSNKNTVHTEDLESK 918

Qy        287 VSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGN-----------LNNMNR 334

Db        919 MKEVEL---KNTELMERID-SLEAERFVASSIEKSRIQKLVNEFDNLKQKLDNDMSNYSK 974

Qy        335 DLAF------HLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPE-LLREMMSEQN-GD 386

Db        975 EKQWLQWRISNLEKDNSELQKQIQPSS------EKSLE----SLNKGTLRKTMSEPDFGD 1024

Qy        387 ALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQ--------------TPQAKPSVSDPR 431

Db       1025 DMSTEGDGASTNESADFMVESVTAPVLSLKSPQPFSQLADVLNLVRSDLEQVLTEIEEPE 1084

Qy        432 SVEEHVSQTSPVDTQSKQDQ 451
```

```
Db       1085 AAKQEEPQMSLEKSTNKNVQ 1104
```

RESULT 8
E98012
hypothetical protein smc [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C;Accession: E98012
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: E98012
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1179 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAK99929.1; PID:g15458753; GSPDB:GN00174
C;Genetics:
A;Gene: smc

  Query Match 5.7%; Score 134.5; DB 2; Length 1179;
  Best Local Similarity 21.6%; Pred. No. 1.1;
  Matches 100; Conservative 80; Mismatches 178; Indels 105; Gaps 24;

```
Qy          5 KHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEA 64

Db        674 KPELEQLQKEIAADE---ASLGSEEAALKTLQDQMAALTERLEAIK---SQGEQARIQEQ 727

Qy         65 CLNKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDH 124

Db        728 GLS-------LAYQQTSQQVEELETL-WKLQEE-EIDRLSEG-----DWQAD-KEKCQES 772

Qy        125 L--------NPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPL 176

Db        773 LATIASEKQNLEAEIEEIKSNKNAIQERYQ-------------NLQEEVAQARLLKTEL 818

Qy        177 VGEVTFFDFQSAERKGEISQLKSMLTTTVAQERF--AIQFKMENAKRCVTQLDRLSALVS 234

Db        819 QGQKR-YEVADIERLG-----KELDNLNIEQEEIQRMLQEKVDNLEKVDTE------LLS 866

Qy        235 TKCHSLGSQSTNF--GFAKSL--LTRVENALVHLSG------------IKLAPKAEAKTV 278

Db        867 QQAEESKTQKTNLQQGLIRKQFELDDIEGQLDDIASHLDQARQQNEEWIRKQTRAEAK-- 924

Qy        279 EQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLS--ELGNLNNMNRDL 336

Db        925 KEKVSERL-------RHLQNQLTDQYQISYTEALEKAHELENLNLAEQEVQDLEKAIRSL 977

Qy        337 AFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDA---LSTIFN 393

Db        978 GPVNLEAIDQY---EEVHNRLDFL--NSQRDDILSAKNLLLETITEMNDEVKERFKSTFE 1032

Qy        394 AA------------GLNHLDQVLLPEVSTPTVGIESPQTPQAK 424

Db       1033 AIRESFKVTFKQMFGGGQAD-LILTEGDLLTAGVEISVQPPGK 1074
```

RESULT 9
T18302
apsB protein - Emericella nidulans
C;Species: Emericella nidulans, Aspergillus nidulans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000
C;Accession: T18302
R;Suelmann, R.; Sievers, N.; Galetzka, D.; Robertson, L.; Timberlake, W.E.; Fischer, R.
Mol. Microbiol. 30, 831-842, 1998
A;Title: Increased nuclear traffic chaos in hyphae of Aspergillus nidulans: molecular ch
A;Reference number: Z18870; MUID:99140441; PMID:10094631
A;Accession: T18302
A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA
A;Residues: 1-1051 <SUE>
A;Cross-references: EMBL:AJ003163; NID:e1286086; PID:e1286087; PIDN:CAA05918.1
C;Genetics:
A;Gene: apsB
A;Introns: 160/3

Query Match 5.7%; Score 134; DB 2; Length 1051;
Best Local Similarity 19.2%; Pred. No. 1;
Matches 86; Conservative 62; Mismatches 134; Indels 166; Gaps 17;

```
Qy          7 QIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQ---- 62
Db        416 KIEECQR-LTED------LRTQDDNLRALQAEMRSASEGIIRLEE----DAQNNLQRYKA 464

Qy         63 -----EACLNKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADL 117
Db        465 VQQELEQCNEEMESLEKSLYEANSKVQR-------------LTVQIESSQNEIAFLRE- 509

Qy        118 SEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLV 177
Db        510 -------------------------EQDGDKIK-----IGDLESELKTYRMSLQ---- 533

Qy        178 GEVTFFDFQSAERKGEISQLKSMLTTTVAQERFAIQF----KMENAKRCVTQLDRLSALV 233
Db        534 -----------SEKDKTKELEGRL----AEERYQREVVGSKEKQEVQRIMNELNREVSAA 578

Qy        234 STKCH----SLGSQSTNFGFAKSLLTRVENALVHLSG-------------IKLAPKAEAK 276
Db        579 KEECRKLKKNLSAQEIETNTWKERLTDLENNLRETLGDLTGSRSSLIANIMKLQKELEST 638

Qy        277 TVEQEVAESSVSEGE---------LPSH-MDTKHIERIPMASEQA---------QTVSQH 317
Db        639 ALELESTRSTLDEKETLLRNRDALLESHGLESRKLSELLERERQARRADKQSFEQALKSH 698

Qy        318 LHA--------GNLSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGY 369
Db        699 HQASRTITQNNSRILELENARNQDRK-------------------------------RF 726

Qy        370 LSLPELLREMMSEQNGDALSTIFNAAGL 397
Db        727 TSLEQQFREQLNERNSMLLTIWKRLSGM 754
```

RESULT 10
T22976
hypothetical protein F59A2.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T22976; T23157
R;Lightning, J.
submitted to the EMBL Data Library, June 1994
A;Reference number: Z19645
A;Accession: T22976
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1133 <WIL>
A;Cross-references: EMBL:Z34801; PIDN:CAA84332.1; GSPDB:GN00021; CESP:F59A2.6
A;Experimental source: clone F59A2
R;Burton, J.
submitted to the EMBL Data Library, October 1995
A;Reference number: Z19700
A;Accession: T23157
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1133 <WI2>
A;Cross-references: EMBL:Z66514; PIDN:CAA91344.1; GSPDB:GN00021; CESP:F59A2.6
A;Experimental source: clone K01A11
C;Genetics:
A;Gene: CESP:F59A2.6
A;Map position: 3
A;Introns: 13/2; 43/3; 107/3; 413/3; 492/3; 567/3; 635/3; 710/3; 738/3; 795/2; 1008/3; 1

Query Match 5.5%; Score 129.5; DB 2; Length 1133;
Best Local Similarity 21.3%; Pred. No. 2.2;
Matches 110; Conservative 77; Mismatches 184; Indels 145; Gaps 23;

```
Qy          9 EQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADE----RDALQEA 64
Db        116 EQLAK----------------AMEKLNSEQNILDEVTKKLEQ--SEEEVLAARGAIQE- 155

Qy         65 CLNKWKILSDSLYEQFSKTT---RDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKH 121
Db        156 -------LTEKLEESEKETSTAKTELEAVS----------KKLDSSETSLKEFSDMIE-- 196

Qy        122 WDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLVGEVT 181
Db        197 ------AMKIQLINCEKQKDEAVELLKQKLEEVEKNMSDVE---------VQKQLLLEST 241

Qy        182 FFDF------------QSAERKGEISQLK-------SMLTTTVAQERFAIQ---FKMENA 219
Db        242 TSEMKQHAEAAEIVKKQLEEAQSSIENLKKDAENERNLKTALESDESSAISEITKQMEAA 301

Qy        220 KRCV-----------TQLDRLSALVSTKCHSLGSQ-------STNFGFAKSLLTRVENAL 261
Db        302 KKELEASEKEKSELREQMDRLQ-----KVHNAGQEDIQKLQKTWELEMAKIAKSTEDEKL 356

Qy        262 V--HLSGIKLAPKAEAKTVEQEV------AESSVSEGELPSHMDTKHIERIPMASEQAQT 313
Db        357 AREQLAGELENAKEDLKVVEEEKHTGIQRAQGALDDAEKEVKVLKEQLERAQSALESSQE 416

Qy        314 VSQHLHAGNLSEL-GNLNNMNR------DLAFHLLREVSDYFRQSEPHSPISFLLEKAIR 366
Db        417 LASSQKADKIQELEKELQNAQKRSSEELETANEMVRSLTATLENS--NSETEILKQK--- 471

Qy        367 WGYLSLPELLREMMSEQNGD-ALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKP 425
Db        472 -----LETLDKELQARQQTEKALTEEINVLTTS------LAEKEQQTAQIQNLQTQIYQM 520

Qy        426 SVSDPRSVEEHVSQTSPVDTQSKQDQKPQSSATSAL 461
Db        521 EVEKEEKVE-------LVKVQLQQAAQSSSSAEEAL 549
```

RESULT 11
T42977
large tegument protein - ateline herpesvirus 3 (strain 73)
C;Species: ateline herpesvirus 3
A;Variety: strain 73
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C;Accession: T42977
R;Albrecht, J.C.; Fleckenstein, B.
submitted to the EMBL Data Library, August 1998
A;Description: Primary structure of the herpesvirus ateles genome.
A;Reference number: Z22274
A;Accession: T42977
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2471 <ALB>
A;Cross-references: EMBL:AF083424; PIDN:AAC95588.1
A;Experimental source: strain 73

Query Match 5.4%; Score 127.5; DB 2; Length 2471;
Best Local Similarity 19.1%; Pred. No. 9.2;
Matches 121; Conservative 90; Mismatches 187; Indels 237; Gaps 31;

```
Qy          1 MPLSKHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQT-ALRKLSQN-----PS 54
Db        660 LPTKVLEIQPLEK--:----SKVFLEKLKPVEQKLTTEVNELLTDLLHNIKQDATEILPV 712

Qy         55 ADERDALQ---------EACLNKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLE 105
Db        713 PDFTTILKNIQSTLQLLHTCVTDLNI--DK--EVIGSTIQQLSYIGWEV---------- 757

Qy        106 SAANSLEWLADLSEKHWDH--LNPVLPVETL---KSDDDKGKEREQADAKVKAFF----Q 156
```

```
Db        758 ---------AELSHKQWNFPPASPIIPLQLLDEIKTEVQRVTAKQQTQETLNQILSDVQS 808

Qy        157 LVGDSEESSILYAPVLQL------PLVGE---VTFFDFQSAERKGEISQ--LKSMLTTT- 204
              |:  :|:|| |  |:||       |||    ||   :   :|   |:  ||::: :|
Db        809 LLSQAEQSSTLSIPILQHYITQAGTLVGESKNETFESLRDTVQKLSTSEEFLKNLINSTT 868

Qy        205 ------------------------------------------------------------V 205
                                                                           :
Db        869 LLNIQTQIKDISDILLSNQYMYQSETIKHAFLEKSNTLIGEAIQLINLKKYTDLTQPILI 928

Qy        206 AQERFAIQFK------------------------MENAKRCVTQLDRLSALVSTKCHS 239
              | :||  : |                      :|  |  :  :| |   ::|
Db        929 AVKRFLSEAKFKDSDNIYEIIYTLTSIGSLLSNSPTIEALKDTLKSIDTLKPKLATVSKP 988

Qy        240 LGSQSTNF--GFAKSLLTRVE---------------------NALVHLSGIKLAPKAEA 275
              |  :  |      | | |:::                    ||      |: ||  :|
Db        989 LKRELYNVMRKLQKDLKTQLQKQELENWKMEVETFASTPGRDVNAF-----IQSAPSTKA 1043

Qy        276 KTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRD 335
              :  :|| :  : |      ||     | :: |   :  :: |        : :  :|
Db       1044 RKYAKEVLKDQIQE------MD------IDVSPE--SIIDDNIKANGQKAWKKIQSAFQD 1089

Qy        336 LAFHLLREVSDYFRQSEPHS-PISFL-------LEKAIRWGYLSLPEL----LREMMSEQ 383
              | | :|   |:   :: :| | | |       | | :  ::|:  :    |: ::
Db       1090 LNFSIL-TADDWLSLAKEYSRPDSSLFTTIGPTLLKLVEEVFVSIQNIKDAKLKSLL--P 1146

Qy        384 NGDALST------IFNAAGLN-HLDQVLLPEVST--PTVGIESPQTPQAKPSVSDPRSV- 433
              || | ::       |    :| ||  : ||::||   |:| |     ||  | : | :|
Db       1147 NGPAFTSPKLDWIHFYEDNVNFHLKTINLPKISTIAHTIGHELSMLSQALNSKTLPEAVV 1206

Qy        434 ----EEHVSQTS-------------PVDTQSKQDQ 451
                  |:| :: |               |||::| |:
Db       1207 GTPLEQHAAKFSCMFKTLEATWHDHQVDTRTKIDE 1241
```

RESULT 12
AE1643
ATP-dependent dsDNA exonuclease SbcC homolog sbcC [imported] - Listeria innocua (strain
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AE1643
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AE1643
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1023 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC96917.1; PID:g16414173; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: sbcC

  Query Match   5.4%; Score 126.5; DB 2; Length 1023;
  Best Local Similarity   20.3%; Pred. No. 3;
  Matches 102; Conservative   82; Mismatches 192; Indels 127; Gaps   21;

```
Qy          8 IEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACLN 67
              ||:::|           |: ||  :: : |     |: |     ||:           |  :|
Db        284 IEEVAK-----------LEAEKDFYQVIANRIEAAKRASNLRSQDALCIRLKEQLETAVN 332

Qy         68 KWKILS---DSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEK---- 120
                | ::   : :  |||   :  | :    |  |   : ||:   :|  | ::  |
Db        333 TEKQVAHEVELITNQFSNAKKQKEAL----AEQ---EAVLEANKRTLFQLEEMEPKIIEL 385

Qy        121 ------------HW----DHLNPVLPVE-----TLKSDDDKGKEREQADAKVKAFFQLVG 159
                         :|    | |  :  |      :| : :  |  ||:    |  : :
Db        386 ETVTIQKRRAEINWKEAADRLEKVMKTEQEIAAELQSMETRLGEINQAEL---AILEAIN 442

Qy        160 DSEESSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSMLTTTVAQERFAIQ--FKME 217
                  :  |          || :    |  : ::: |   || :||      |: ||:   | |
Db        443 KRTTAEALIEK--NQELVNKRMKMDLWNKQKQTEEQTLKQLLT-----EKLAIETMIKQE 495

Qy        218 NAKRCVTQLDRLSALV--STKCHSLGS----QSTNFGFAKSLLTRVENALVHLSGIKLAP 271
               :     |   |:| :    |   ||     |   :| : || | |
Db        496 ESNLQQEQAATLAAHLHEGDACPVCGSVSHPQLAEYGESASLTTLEE------------ 542

Qy        272 KAEAKTVEQEVA----ESSVS--EGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSE 325
               |:||  |:::|    | |:|  | :|    |   :: : :    |:
Db        543 -AKAKLHEKQLAINEVEKSISQLEWQLAEWADIADVDLVAVEKLLAE------------ 588

Qy        326 LGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNG 385
                    || || :|  :::          |   ||         :|    :|: :|:|
Db        589 -------NRQLAKNLTDQINQLQTNVAQKENIQGALE--------TLKNKQKELETEKNN 633

Qy        386 DAL-------STIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPRSVEEHVS 438
               ||            ::  |::|:| :  ::   |:        : |  :|:  |:: |:
Db        634 TALQVESFHQEVQLSSGKLSYLEQAIPADLRDKTI------FDKKKNELSN--SIKTHLE 685

Qy        439 QTSPVDTQSKQDQKPQSSATSAL 461
               |   ||   :: :|  :   | |
Db        686 QAEQVDKIFREAEKETTRLESTL 708
```

RESULT 13
S50914
cell division control protein CDC37 - yeast (Saccharomyces cerevisiae)
N;Alternate names: CDC37 start control protein; protein YD9489.03; protein YDR168w
C;Species: Saccharomyces cerevisiae
C;Date: 10-Feb-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C;Accession: S50914; B26372
R;Oliver, K.; Harris, D.
submitted to the EMBL Data Library, January 1994
A;Reference number: S50912
A;Accession: S50914
A;Molecule type: DNA
A;Residues: 1-506 <OLI>
A;Cross-references: EMBL:Z47813; NID:g642274; PIDN:CAA87799.1; PID:g642277; MIPS:YDR168w
R;Ferguson, J.; Ho, J.Y.; Peterson, T.A.; Reed, S.I.
Nucleic Acids Res. 14, 6681-6697, 1986
A;Title: Nucleotide sequence of the yeast cell division cycle start genes CDC28, CDC36,
A;Reference number: A93635; MUID:86312926; PMID:3018676
A;Accession: B26372
A;Molecule type: DNA
A;Residues: 58-168,'D',170-506 <FER>
A;Cross-references: EMBL:X04288; NID:g3492; PIDN:CAA27836.1; PID:g3493
A;Note: the authors translated the codon GAC for residue 112 as Ala
C;Genetics:
A;Gene: SGD:CDC37
A;Cross-references: SGD:S0002575; MIPS:YDR168w
A;Map position: 4R
C;Keywords: cell cycle control

  Query Match   5.4%; Score 126; DB 2; Length 506;
  Best Local Similarity   19.1%; Pred. No. 1.2;
  Matches   90; Conservative   71; Mismatches 161; Indels 148; Gaps   17;

```
Qy         15 LSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACLNKWKILSD 74
              |||||   |:  ::| :|                                    |||
Db         13 LSDDSDVEVHPNVDKKSF---------------------------------IKWK--QQ 36

Qy         75 SLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNPVLPVETL 134
              |:||  |  :||: :   |      |:  ::        |::| |      |  |
Db         37 SIHEQRFKRNQDIKNLETQVDMYSHLNKRVDRI------LSNLPESSLTDLPAVTKFLNA 90

Qy        135 KSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAERKGEI 194
                 |  :  : |  | ::  : ::| | |  |              ||           :  :::|:
Db         91 NFDKMEKSKGENVDPEIATYNEMVEDLFE---------QLA----------KDLDKEGKD 131

Qy        195 SQLKSMLTTTVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLL 254
              |:  |::   : : |  |      ||:      :|  |   |   : |:  :  ||  | :
```

```
Db        132 SKSPSLIRDAILKHRAKIDSVTVEAKK------KLDELYKEKNAHISSEDIHTGFDSSFM 185

Qy        255 TRVENALVHLSGI---KLAPKAEAKTVEQEVAESSVSE------------------GEL 292

Db        186 NKQKGGAKPLEATPSEALSSAAES-NILNKLAKSSVPQTFIDFKDDPMKLAKETEEFGKI 244

Qy        293 PSHMDTKH----IERIPMASEQ------AQTVSQHLHAG---------------------- 321

Db        245 SINEYSKSQKFLLEHLPIISEQQKDALMMKAFEYQLHGDDKMTLQVIHQSELMAYIKEIY 304

Qy        322 --------NLSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLP 373

Db        305 DMKKIPYLNPMELSNVINM-------FFEKV--IFNKDKPMGKESFL--RSVQEKFLHIQ 353

Qy        374 E----LLREMMSEQNGDALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQ 419

Db        354 KRSKILQQEEMDESNAEGVETI----QLKSLDDSTELEVNLPDFNSKDPE 399


RESULT 14
T22235
hypothetical protein F45G2.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T22235
R;Lindsay, S.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19535
A;Accession: T22235
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1287 <WIL>
A;Cross-references: EMBL:Z93382; PIDN:CAB07612.1; GSPDB:GN00021; CESP:F45G2.3
A;Experimental source: clone F45G2
C;Genetics:
A;Gene: CESP:F45G2.3
A;Map position: 3
A;Introns: 59/2; 117/3; 153/3; 180/2; 233/2; 341/3; 393/1; 424/2; 551/3; 664/3; 734/3; 8

  Query Match             5.4%;  Score 125.5;  DB 2;  Length 1287;
  Best Local Similarity   21.4%;  Pred. No. 4.9;
  Matches   90;  Conservative   79;  Mismatches  167;  Indels   85;  Gaps   18;

Qy          7 QIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACL 66

Db        259 EIEELNRNLKAESESNYENAKKASELERLR------EKEKKEWDEKERRMQMEAENEA-- 310

Qy         67 NKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLN 126

Db        311 NQHKTQTEKLKDQISTLQTDFEKVSAQRKAQEQMNAELVDEVASFKQKAERAEEQKKKL- 369

Qy        127 PVLPVETLKSDDDK-GKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPL-VGEVTFFD 184

Db        370 ----VEDLDSLDDKLAAEKRANNEHVKHNKKLEGQ------LKATQTQLTLALKEKHEFD 419

Qy        185 FQSAERKGEISQLKSMLTTTVAQERFAIQFKME-NAKRCVTQLDRLSALVSTKCHSLGSQ 243

Db        420 VDCKRRESEIGELK-----LKAQGDANLISKLQAMLRKCISRIEELEEDLLEE------- 467

Qy        244 STNFGFAKSLLTRVENALVHLSGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTKHIER 303

Db        468 -------RKLRMKAERQFNEL-------RSEYEVLQEQMAEAS---GQLTAE---AHINK 507

Qy        304 IPMASEQAQTVSQHLHAGNLSE--------------LGNLNNMNR---DLAFHLLREVSD 346

Db        508 V--RAEEVSNLRRDLQKRNLNHEAYISDLCNMQYATVNNLRNLSQQFYSFNFNCFFFLFL 565

Qy        347 YFRQSEPHSPIS-FLLEKAIRWGYLSLPELLREMMSEQNGDALST--------IFNAAGL 397

Db        566 FFSFSIYDKPIHLFQIRKIIK---SPVTVIFHELISRESVPTVASLCIYAQTIIFNGASL 622

Qy        398 N 398

Db        623 N 623


RESULT 15
A47297
myosin heavy chain form B, nonmuscle - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 02-Feb-2001
C;Accession: A47297; A55441
R;Bhatia-Dey, N.; Adelstein, R.S.; Dawid, I.B.
Proc. Natl. Acad. Sci. U.S.A. 90, 2856-2859, 1993
A;Title: Cloning of the cDNA encoding a myosin heavy chain B isoform of Xenopus nonmusc
A;Reference number: A47297; MUID:93219383; PMID:8464900
A;Accession: A47297
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1992 <BHA>
A;Cross-references: GB:L09740; NID:g214623; PIDN:AAA49915.1; PID:g214624
A;Experimental source: XTC cells
A;Note: sequence extracted from NCBI backbone (NCBIP:128722)
R;Kelley, C.A.; Oberman, F.; Yisraeli, J.K.; Adelstein, R.S.
J. Biol. Chem. 270, 1395-1401, 1995
A;Title: A Xenopus nonmuscle myosin heavy chain isoform is phosphorylated by cyclin-p34
A;Reference number: A55441; MUID:95138137; PMID:7836406
A;Accession: A55441
A;Status: preliminary
A;Molecule type: protein
A;Residues: 198-232 <KEL>
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: nucleotide binding; P-loop
F;88-787/Domain: myosin motor domain homology <MMOT>
F;178-185/Region: nucleotide-binding motif A (P-loop)

  Query Match             5.4%;  Score 125.5;  DB 2;  Length 1992;
  Best Local Similarity   21.2%;  Pred. No. 9.1;
  Matches  107;  Conservative   78;  Mismatches  189;  Indels  131;  Gaps   20;

Qy          1 MPLSKHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDA 60

Db        929 LAIKKQEMEEILRDLE------IRMEEEEERNQVLQNEKKKMQTHVQDLEE--QLDEEEA 980

Qy         61 LQEACLNK-------WKILSDSLY--EQFSKTTRDIELISWFVAAQFLLDTTLESAANSL 111

Db        981 AQKLQLEKVTAEAKIKKMEEDILVLEDQNSKFLKEKKLLEERIAESTSQLAEEEEKAKNL 1040

Qy        112 EWLADLSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPV 171

Db       1041 AKLKNKQEMMISDLE-----ERLKKEEKTRQELEKAKRK---------------------- 1074

Qy        172 LQLPLVGEVTFFDFQSAERKGEISQLKSMLTTTVAQERFAIQFKMENAKRCVTQ------ 225

Db       1075 ----LDGETTDFQDQIAELQAQIEELKLQL----AKKEEELQAALARGDEEVLQKNNTLK 1126

Qy        226 -LDRLSALVSTKCHSLGSQSTNFGFA----------KSLLTRVENALVHLSGIKLAPKA 273

Db       1127 LVRELQAQIAELQEDLESEKASRNKAEKQKRDLSEELEALKTELEDTL-----DTTAAQQ 1181

Qy        274 EAKTV-EQEVAESSVS-EGELPSH---------MDTKHIERIPMASEQAQTVSQHLHAGN 322

Db       1182 ELRTKREQEVAELRKSIEEETRNHEAQIQEMRQRQATALEELSEQLEQAKRFKVNLEKNK 1241

Qy        323 LSELGNLNNMNRDLAFHL-----LREVSDYFRQ------SEPHSPISFLLEKAIRWGYLS 371

Db       1242 QS----LESDNKELATEVKSLQQMKAESEYKRKKLEGQVQELHAKV---LE--------- 1285

Qy        372 LPELLREMMSEQNGDALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPR 431

Db       1286 -GDRLRADMVEKSSKLQNELENVSS-------LLEEAEKKGIKL----------AKDVA 1326

Qy        432 SVEEHVSQTSPVDTQSKQDQKPQSS 456

Db       1327 SMESQLQDTQELLQEETRQKLNQSS 1351
```

Search completed: July 3, 2004, 02:46:49
Job time : 30 secs

GenCore version 5.1.6

CM protein - protein search, using sw model

Run on: July 2, 2004, 22:59:50 ; Search time 19 Seconds
(without alignments)
1268.867 Million cell updates/sec

Title: US-10-780-347-2
Perfect score: 2343
Sequence: 1 MPLSKHQIEQLSKPLSDDSI.....DTQSKQDQKPQSSATSALSW 463

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	152	6.5	2245	1	MYSJ_DICDI	P54697 dictyosteli
2	134	5.7	1051	1	APSB_EMENI	O60039 emericella
3	129.5	5.5	790	1	EEA1_MOUSE	Q8b166 mus musculu
4	128.5	5.5	877	1	CAPP_VIBVU	Q8dcn2 vibrio vuln
5	126	5.4	506	1	CC37_YEAST	P06101 saccharomyc
6	125	5.3	793	1	REGA_DICDI	Q23917 dictyosteli
7	125	5.3	1194	1	SMC3_SCHPO	O42649 schizosacch
8	124.5	5.3	1940	1	MYH3_HUMAN	P11055 homo sapien
9	124.5	5.3	1940	1	MYH3_RAT	P12847 rattus norv
10	124	5.3	866	1	MYSP_SCHMA	P06198 schistosoma
11	123.5	5.3	867	1	DPO1_RICPR	O05949 rickettsia
12	123	5.2	1935	1	MYH7_HUMAN	P12883 homo sapien
13	123	5.2	1935	1	MYH7_RAT	P02564 rattus norv
14	123	5.2	2022	1	ANT1_ONCVO	P21249 onchocerca
15	122.5	5.2	1220	1	LBN_MOUSE	Q8k1g2 mus musculu
16	122.5	5.2	2469	1	TEGU_HSVSA	Q01056 herpesvirus
17	122	5.2	1934	1	MYH7_MESAU	P13540 mesocricetu
18	121.5	5.2	1679	1	GCC2_MOUSE	Q8chg3 mus musculu
19	121.5	5.2	3259	1	GOB1_HUMAN	Q14789 homo sapien
20	121	5.2	1357	1	KTN1_HUMAN	Q86up2 homo sapien
21	120.5	5.1	999	1	OXRP_HUMAN	Q9y411 homo sapien
22	120.5	5.1	1938	1	MYH4_RABIT	Q28641 oryctolagus
23	120	5.1	667	1	MI10_CAEEL	P34400 caenorhabdi
24	119.5	5.1	1066	1	KL61_DROME	P46863 drosophila
25	119	5.1	1935	1	MYH7_PIG	P79293 sus scrofa
26	119	5.1	2104	1	MYS3_SCHPO	O14157 schizosacch
27	118.5	5.1	1920	1	PCT2_MOUSE	P48725 mus musculu
28	118.5	5.1	1937	1	MYH8_HUMAN	P13535 homo sapien
29	118	5.0	1979	1	TRIA_HUMAN	Q15643 homo sapien
30	117.5	5.0	539	1	YAH2_YEAST	P39705 saccharomyc
31	117.5	5.0	1427	1	REST_HUMAN	P30622 homo sapien
32	117.5	5.0	2116	1	MYS2_DICDI	P08799 dictyosteli
33	117.5	5.0	2611	1	BP1E_MOUSE	Q91zu8 mus musculu
34	117	5.0	1679	1	YIO9_YEAST	P40457 saccharomyc
35	117	5.0	2779	1	LVA_DROME	Q8mss1 drosophila
36	116	5.0	1583	1	GCC2_HUMAN	Q8iwj2 homo sapien
37	116	5.0	1616	1	P200_MYCGE	Q49429 mycoplasma
38	116	5.0	1966	1	MYSB_CAEEL	P02566 caenorhabdi
39	115.5	4.9	1690	1	C190_DROME	Q9vje5 drosophila
40	115.5	4.9	2468	1	MAPB_HUMAN	P46821 homo sapien
41	115.5	4.9	3660	1	DMD_CHICK	P11533 gallus gall
42	115	4.9	1312	1	RA50_YEAST	P12753 saccharomyc
43	114.5	4.9	955	1	KINL_LEICH	P46865 leishmania
44	114.5	4.9	2230	1	GOA4_HUMAN	Q13439 homo sapien
45	114.5	4.9	5327	1	MACF_MOUSE	Q9qxz0 mus musculu

ALIGNMENTS

RESULT 1
MYSJ_DICDI
ID MYSJ_DICDI STANDARD; PRT; 2245 AA.
AC P54697;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin IJ heavy chain.
GN MYOJ.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=96215148; PubMed=8636147;
RA Hammer J.A. III, Jung G.;
RT "The sequence of the dictyostelium myo J heavy chain gene predicts a
RT novel, dimeric, unconventional myosin with a heavy chain molecular
RT mass of 258 kDa.";
RL J. Biol. Chem. 271:7120-7127(1996).
RN [2]
RP SEQUENCE OF 1-1021 FROM N.A.
RX MEDLINE=97039016; PubMed=8884597;
RA Peterson M.D., Urioste A.S., Titus M.A.;
RT "Dictyostelium discoideum myoJ: a member of a broadly defined myosin
RT V class or a class XI unconventional myosin?";
RL J. Muscle Res. Cell Motil. 17:411-424(1996).
RN [3]
RP SEQUENCE OF 182-298 FROM N.A.
RX MEDLINE=95023928; PubMed=7937787;
RA Titus M.A., Kuspa A., Loomis W.F.;
RT "Discovery of myosin genes by physical mapping in Dictyostelium.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9446-9450(1994).
CC -!- FUNCTION: Myosin is a protein that binds to actin and has ATPase
CC activity that is activated by actin.
CC -!- SUBUNIT: Homodimer.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 dilute domain.
CC -!- SIMILARITY: Contains 3 IQ domains.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; U42409; AAA85186.1; -.
DR EMBL; L35322; AAA79858.1; -.
DR PIR; T18278; T18278.
DR HSSP; P08799; 1MND.
DR DictyBase; DDB0185050; myoJ.
DR InterPro; IPR002710; DIL.
DR InterPro; IPR000048; IQ_region.

```
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR004009; Myosin_N.
DR   Pfam; PF01843; DIL; 1.
DR   Pfam; PF00612; IQ; 6.
DR   Pfam; PF00063; myosin_head; 2.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD003376; DIL; 1.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 3.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 3.
KW   Myosin; ATP-binding; Actin-binding; Calmodulin-binding; Repeat;
KW   Coiled coil.
FT   DOMAIN        1    809       MYOSIN HEAD-LIKE.
FT   DOMAIN      824    851       IQ 1.
FT   DOMAIN      872    901       IQ 2.
FT   DOMAIN      943    972       IQ 3.
FT   DOMAIN      973   1812       COILED COIL (POTENTIAL).
FT   DOMAIN     2060   2220       DILUTE.
FT   NP_BIND     174    181       ATP (POTENTIAL).
FT   DOMAIN      669    749       ACTIN-BINDING.
FT   CONFLICT    191    191       L -> F (IN REF. 2).
FT   CONFLICT    284    284       A -> T (IN REF. 2).
FT   CONFLICT    291    291       G -> R (IN REF. 2).
FT   CONFLICT    332    347       NKSGCFEIEGVSDEEH -> IEWMFELKVYRMKS (IN
FT                                REF. 2).
FT   CONFLICT    550    550       N -> K (IN REF. 2).
FT   CONFLICT    865    866       HH -> QQ (IN REF. 2).
SQ   SEQUENCE   2245 AA;  258478 MW;  615E5EF1D1AB45BE CRC64;

  Query Match             6.5%;  Score 152;  DB 1;  Length 2245;
  Best Local Similarity   23.6%;  Pred. No. 0.19;
  Matches   81;  Conservative   57;  Mismatches  143;  Indels   62;  Gaps   13;

Qy         23 VYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACLNKWKILSDSLYEQFSK 82
              : |: |  : | :: : |  |  | :|    ::: :   |    |   |: ||:   : |
Db        964 IQLRAEARSLRTVQEQKNKLQEKLEELQWRLTSEAKRKQQ---LEDQKVKSDTTISELSS 1020

Qy         83 TTRDIEL-ISWFVAAQFLLDTTLESAANSL-EWLADLSEK-----HWDHLNPVLPVETLK 135
                :|| :|        || : :|:   | | |: | |:      |   || |  :
Db       1021 NNDHLELQLSEIQLKYQELDNSNQSSQLQLSECLSKLEEQTQQLDHSSKLNKKLEKDLSD 1080

Qy        136 SDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLVGEVTF----FDFQSAER- 190
               |  | :  |  :   :   |   ||| |   :   |     :: |    ||  | ||
Db       1081 QHDSIEKLQSQFNETEQQLQQFKQQSEELSSKLSKTTQ-----QLDFNKQEFDRLSQERD 1135

Qy        191 ------KGEISQLKSMLTTTVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQS 244
                  : || |||    :|: :: |::    :| :| | :|   : |:  :  ||| ||
Db       1136 TDNTNNQLEIQQLKK-ANSTLEEDYFSLSGIRDNLERQVLELRDENQLIKERLDSLGQQS 1194

Qy        245 TNF--GFA------KSLLTRVENALVHLSGIKLAPKAEAK-----------------TVE 279
              : |  | |      : |:      |: ||  ||  : |||                  ::
Db       1195 SQFQSGAALEKQQLEQLVQEQSEQLIKLSSEKLGSEEEAKKQINQLELELTDHKSKLQIQ 1254

Qy        280 QEVAESSVS-----EGELPSHMDTK-----HIERIPMASEQAQ 312
               :: | |       :|:|  : | |      :|||  :  :  :
Db       1255 LQLTEQSNEKIKKLKGKLEEYQDEKKQLQQELERIKQSKQSVE 1297


RESULT 2
APSB_EMENI
ID   APSB_EMENI     STANDARD;      PRT;  1051 AA.
AC   O60039;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Anucleate primary sterigmata protein B.
GN   APSB.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; Emericella.
OX   NCBI_TaxID=162425;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RX   MEDLINE=99140441; PubMed=10094631;
RA   Suelmann R., Sievers N., Galetzka D., Robertson L., Timberlake W.E.,
RA   Fischer R.;
RT   "Increased nuclear traffic chaos in hyphae of Aspergillus nidulans:
RT   molecular characterization of apsB and in vivo observation of nuclear
RT   behaviour.";
RL   Mol. Microbiol. 30:831-842(1998).
CC   -!- FUNCTION: Involved in regulation of nuclear migration. May be
CC       involved in regulating nuclear positioning.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: SOME, TO TRICHOHYALIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AJ003163; CAA05918.1; -.
DR   PIR; T18302; T18302.
KW   Coiled coil.
FT   DOMAIN       10    200       COILED COIL (POTENTIAL).
FT   DOMAIN      239    285       COILED COIL (POTENTIAL).
FT   DOMAIN      325    743       COILED COIL (POTENTIAL).
FT   DOMAIN      787    878       COILED COIL (POTENTIAL).
FT   DOMAIN      950   1004       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1051 AA;  120996 MW;  9262C1A9E2D8847A CRC64;

  Query Match             5.7%;  Score 134;  DB 1;  Length 1051;
  Best Local Similarity   19.2%;  Pred. No. 0.91;
  Matches   86;  Conservative   62;  Mismatches  134;  Indels  166;  Gaps   17;

Qy          7 QIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQ---- 62
              :||:  : |::|       |: :    | |: |    |   : :| :    | :: ||
Db        416 KIEECQR-LTED------LRTQDDNLRALQAEMRSASEGIIRLEE----DAQNNLQRYKA 464

Qy         63 -----EACLNKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADL 117
                  | |  : : |  |||| || |                |   :||: | : :| :
Db        465 VQQELEQCNEEMESLEKSLYEANSKVQR-------------LTVQIESSQNEIAFLRE- 509

Qy        118 SEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLV 177
                                          ||   |:|    :|| |     |   ||
Db        510 --------------------------EQDGDKIK-----IGDLESELKTYRMSLQ---- 533

Qy        178 GEVTFFDFQSAERKGEISQLKSMLTTTVAQERFAIQF----KMENAKRCVTQLDRLSALV 233
                      | :  :|:  |    |:||:  :     :  :|  : :|:|  :
Db        534 -----------SEKDKTKELEGRL----AEERYQREVVGSKEKQEVQRIMNELNREVSAA 578

Qy        234 STKCH----SLGSQSTNFGFAKSLLTRVENALVHLSG-------------IKLAPKAEAK 276
               :|      :| :|        |  || :|| |    |                :||  : |:
Db        579 KEECRKLKKNLSAQEIETNTWKERLTDLENNLRETLGDLTGSRSSLIANIMKLQKELEST 638

Qy        277 TVEQEVAESSVSEGE---------LPSH-MDTKHIERIPMASEQA---------QTVSQH 317
               :| |    |:: | |          | || :::: :  :     ||          | :  |
Db        639 ALELESTRSTLDEKETLLRNRDALLESHGLESRKLSELLERERQARRADKQSFEQALKSH 698

Qy        318 LHA--------GNLSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGY 369
                         |          : || |  | :|                       :
Db        699 HQASRTITQNNSRILELENARNQDRK-------------------------------RF 726

Qy        370 LSLPELLREMMSEQNGDALSTIFNAAGL 397
               || :  || ::|:|    |:     :|:
Db        727 TSLEQQFREQLNERNSMLLTIWKRLSGM 754


RESULT 3
EEA1_MOUSE
ID   EEA1_MOUSE     STANDARD;      PRT;   790 AA.
AC   Q8BL66;
```

```
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Early endosome antigen 1 (Fragment).
GN   EEA1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Brain;
RX   MEDLINE=22354683; PubMed=12466851;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
CC   -!- FUNCTION: Binds phospholipid vesicles containing
CC       phosphatidylinositol 3-phosphate and participates in endosomal
CC       trafficking (By similarity).
CC   -!- SUBUNIT: Homodimer. Binds STX6. Binds RAB5A, RAB5B, RAB5C and
CC       RAB22A that have been activated by GTP-binding (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and peripheral membrane protein
CC       associated with early endosomes (By similarity).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AK046231; BAC32647.1; -.
DR   MGD; MGI:2442192; Eea1.
KW   Coiled coil.
FT   DOMAIN        1    789       COILED COIL (POTENTIAL).
FT   DOMAIN      180    550       GLN/GLU/LYS-RICH.
FT   NON_TER     790    790
SQ   SEQUENCE   790 AA;  90539 MW;  A167AE02CC2CA7BF CRC64;

  Query Match             5.5%;  Score 129.5;  DB 1;  Length 790;
  Best Local Similarity   21.0%;  Pred. No. 1.2;
  Matches  104;  Conservative   85;  Mismatches  175;  Indels  131;  Gaps   21;

Qy          5 KHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLS------QNPSA--- 55
              :||:|:  :   :         |::||   ||   |  :  ||::       ||  |
Db        324 QHQLEKSKQQHQEQQ------ALQQSATAKLREAQNDLEQVLRQIGDKDQKIQNLEALLQ 377

Qy         56 ----------DERDAL---------QEACLNKWKILSDSLYEQFSKTTRDIELISWFVAA 96
                        ||: |          : | ||: :  : :| :| :: |  ::  |
Db        378 KGKESVSLLEKEREDLYAKIQAGEGETAVLNQLQEKNHALQQQLTQLTEKLKNQS----- 432

Qy         97 QFLLDTTLESAANSLEWLADLSEKHWDHL----NPVLPVET----LKSDDDKGKER-EQA 147
                 ||   : | | |  ::   ||    : || :||    | |  :: ||:  |
Db        433 --------ESHKQAEENLHDQVQEQKAHLRAAQDRVLSLETSVSELSSQLNESKEKVSQL 484

Qy        148 DAKVKAFFQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSMLTTTVAQ 207
             | ::||  :|:  :|                          | :  : : |:: | |
Db        485 DIQIKAKTELLLSAE----------------------AAKAAQRADLQNHLDTA--- 516

Qy        208 ERFAIQFKMENAKRCVTQLDRLSALVSTK---CHSLGSQSTNFGFAK-SLLTRVENALVH 263
              : |:| | :   :   |||:|:|     |   |  | |   :       ||  :||:   |
Db        517 -QHALQDKQQELNKVSVQLDQLTAKFQEKQEHCIQLESHLKDHKEKHLSLEQKVEDLEGH 575

Qy        264 LSGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVS-QHLHAGN 322
             :  :: |   | |  :::  :|   : :| : :: :: |       || : ||   |   |
Db        576 IKKLE-ADALEVKASKEQALQSLQQQRQLSTDLELRNAELSRELQEQEEVVSCTKLDLQN 634

Qy        323 LSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSE 382
              ||:  | |: :|          :: | :  :    ||  :       ||
Db        635 KSEI--LENIKQTLT----------KKEEENVVLKQEFEKLSQDSKTQHKEL------- 674

Qy        383 QNGDAL-STIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPRSVEEHVSQTS 441
              || : : :     :     || |:||       | :     |||  |::   |:
Db        675 --GDRMQAAVTELTAVKAQKDALLAELST---------TKEKLSKVSD--SLKNSKSEF- 720

Qy        442 PVDTQSKQDQKPQSS 456
                  |::|| :::
Db        721 -----EKENQKGKAA 730


RESULT 4
CAPP_VIBVU
ID   CAPP_VIBVU     STANDARD;      PRT;   877 AA.
AC   Q8DCN2;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phosphoenolpyruvate carboxylase (EC 4.1.1.31) (PEPCase) (PEPC).
GN   PPC OR VV11369.
OS   Vibrio vulnificus.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=672;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CMCP6;
RA   Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA   Choy H.E.;
RT   "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Forms oxaloacetate, a four-carbon dicarboxylic acid
CC       source for the tricarboxylic acid cycle.
CC   -!- CATALYTIC ACTIVITY: Phosphate + oxaloacetate = H(2)O +
CC       phosphoenolpyruvate + CO(2).
CC   -!- PATHWAY: Tricarboxylic acid cycle.
CC   -!- SIMILARITY: Belongs to the PEPCase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE016801; AAO09818.1; -.
DR   HAMAP; MF_00595; -; 1.
DR   InterPro; IPR001449; PEPcase.
DR   Pfam; PF00311; PEPcase; 1.
```

```
DR   PRINTS; PR00150; PEPCARBXLASE.
DR   PROSITE; PS00781; PEPCASE_1; 1.
DR   PROSITE; PS00393; PEPCASE_2; 1.
KW   Lyase; Carbon dioxide fixation; Tricarboxylic acid cycle;
KW   Complete proteome.
FT   ACT_SITE    138    138       BY SIMILARITY.
FT   ACT_SITE    544    544       BY SIMILARITY.
SQ   SEQUENCE   877 AA;  98900 MW;  14CFD7BE65CA779C CRC64;

  Query Match             5.5%;  Score 128.5;  DB 1;  Length 877;
  Best Local Similarity   23.2%;  Pred. No. 1.6;
  Matches  105;  Conservative   67;  Mismatches  129;  Indels  151;  Gaps   28;

Qy          3 LSKHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRK------LSQNPSAD 56
              :|:|    : :|   |:|  :: || ::      |: : || |:|:      |: :|:
Db         91 ISRHCDAHVCEP---DAINTLFAKLGQNGI----NKLDTAQ-AIRELNIELVLTAHPTEI 142

Qy         57 ERDAL------QEACLNKWKILSDSLYEQFSKTTRDIELI---SW---FVAAQFLLDTTL 104
                |  :          ||:| : |||  |::  || : :| :   ||    :  |    | |
Db        143 TRRTMINKLVKINECLSKLE-LSDLSYKERHKTEKRLEQLIAQSWHSDVIRKQ--RPTPL 199

Qy        105 ESA-------ANSL-----EWLADLSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVK 152
              : |        |||      ::| :| ||  |:|:  ||:                 ||:
Db        200 DEAKWGFAVVENSLWEAVPDFLRELDEKLKDYLDQGLPI----------------DAR-- 241

Qy        153 AFFQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSMLTTTVAQERFAI 212
                                          | |  :  :| |         :| || :|   :
Db        242 -------------------------PVHFSSWMGGDRDG-----NPFVTHTVTREVLLL 270

Qy        213 QFKMENAKRCVTQLDRL-SALVSTKCH----SLGSQSTNFGF------AKSLLTRVENAL 261
                : : |   :  :: | | |  |||:     |  :  :  :        ::||:  :  |
Db        271 S-RWKAADLYLKDINELISELSMTKCNDTVRQLAGEDEHEPYRAILKQLRTLLSDTKEIL 329

Qy        262 -VHLSGIKLAPKAEAKTVEQ------------EVAESSVSEGELPSHMDT--------K 299
                ::| ||| ||  ::|||             |   | ::|| |   :||
Db        330 DAKINGQKLAVKAPLQSVEQLWDPLFACYQSLRECGMSMIAEGSL---LDTLRRVKAFGV 386

Qy        300 HIERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISF 359
              |: |: :  |  :      ||  ||||       | |       : ||   ||    |:|
Db        387 HLVRLDIRQESTR------HADVLSEL------TRYLG------IGDYNHWSE-QDKIAF 427

Qy        360 LLEKAIRWGYLSLPELLREMM-SEQNGDALST 390
              |  :         | | |:   |||  : | |
Db        428 LTNELAS----KRPLLPRDWQPSEQVKEVLDT 455


RESULT 5
CC37_YEAST
ID   CC37_YEAST     STANDARD;      PRT;   506 AA.
AC   P06101; Q04132;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hsp90 co-chaperone Cdc37 (Hsp90 chaperone protein kinase-targeting
DE   subunit) (Cell division control protein 37).
GN   CDC37 OR SMO1 OR YDR168W OR YD9489.03.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE OF 58-506 FROM N.A.
RX   MEDLINE=86312926; PubMed=3018676;
RA   Ferguson J., Ho J.-Y., Peterson T.A., Reed S.I.;
RT   "Nucleotide sequence of the yeast cell division cycle start genes
RT   CDC28, CDC36, CDC37, and CDC39, and a structural analysis of the
RT   predicted products.";
RL   Nucleic Acids Res. 14:6681-6697(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Oliver K., Harris D., Barrell B.G., Rajandream M.A.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   INTERACTION WITH CDC28.
RX   MEDLINE=95273421; PubMed=7753858;
RA   Gerber M.R., Farrell A., Deshaies R.J., Herskowitz I., Morgan D.O.;
RT   "Cdc37 is required for association of the protein kinase Cdc28 with G1
RT   and mitotic cyclins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:4651-4655(1995).
RN   [4]
RP   INTERACTION WITH STE11.
RX   MEDLINE=20130028; PubMed=10664467;
RA   Abbas-Terki T., Donze O., Picard D.;
RT   "The molecular chaperone Cdc37 is required for Ste11 function and
RT   pheromone-induced cell cycle arrest.";
RL   FEBS Lett. 467:111-116(2000).
RN   [5]
RP   INTERACTION WITH CDC28 AND CAK1.
RX   MEDLINE=20094952; PubMed=10629030;
RA   Farrell A., Morgan D.O.;
RT   "Cdc37 promotes the stability of protein kinases Cdc28 and Cak1.";
RL   Mol. Cell. Biol. 20:749-754(2000).
CC   -!- FUNCTION: Co-chaperone that binds to numerous kinases
CC       and promotes their interaction with the Hsp90 complex,
CC       resulting in stabilization and promotion of their activity.
CC   -!- SUBUNIT: Forms a complex with Hsp90. Interacts with CDC28, CAK1
CC       and STE11.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the CDC37 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X04288; CAA27836.1; -.
DR   EMBL; Z47813; CAA87799.1; -.
DR   PIR; S50914; S50914.
DR   GermOnline; 140659; -.
DR   SGD; S0002575; CDC37.
DR   GO; GO:0003754; F:chaperone activity; IDA.
DR   GO; GO:0000074; P:regulation of cell cycle; IMP.
DR   GO; GO:0007103; P:spindle pole body duplication (sensu Saccha. . .; IGI.
DR   InterPro; IPR004918; Cdc37.
DR   Pfam; PF03234; Cdc37; 1.
KW   Chaperone; Cell division; Cell cycle.
FT   CONFLICT    169    169       A -> D (IN REF. 1).
SQ   SEQUENCE   506 AA;  58385 MW;  0DF0C923158A2526 CRC64;

  Query Match             5.4%;  Score 126;  DB 1;  Length 506;
  Best Local Similarity   19.1%;  Pred. No. 1.1;
  Matches   90;  Conservative   71;  Mismatches  161;  Indels  148;  Gaps   17;

Qy         15 LSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACLNKWKILSD 74
              |||||   |:  ::| :|                                      |||
Db         13 LSDDSDVEVHPNVDKKSF---------------------------------IKWK--QQ 36

Qy         75 SLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNPVLPVETL 134
              |::||  |  :||: :   |      |:  ::        |::| |      |  |
Db         37 SIHEQRFKRNQDIKNLETQVDMYSHLNKRVDRI------LSNLPESSLTDLPAVTKFLNA 90

Qy        135 KSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAERKGEI 194
                |  :  : |  | ::  : ::| | |  |          ||           :  :::|:
Db         91 NFDKMEKSKGENVDPEIATYNEMVEDLFE---------QLA----------KDLDKEGKD 131

Qy        195 SQLKSMLTTTVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLL 254
              |:  |::   : : |  |       ||:      :|  |    |   : |:  : ||  | :
Db        132 SKSPSLIRDAILKHRAKIDSVTVEAKK------KLDELYKEKNAHISSEDIHTGFDSSFM 185

Qy        255 TRVENALVHLSGI---KLAPKAEAKTVEQEVAESSVSE------------------GEL 292
                 : :      |       |:  ||:  :  ::|:||| :                 |::
Db        186 NKQKGGAKPLEATPSEALSSAAES-NILNKLAKSSVPQTFIDFKDDPMKLAKETEEFGKI 244
```

```
Qy        293 PSHMDTKH----IERIPMASEQ------AQTVSQHLHAG---------------------- 321
              :  :|     :| :|: |||       :      ||
Db        245 SINEYSKSQKFLLEHLPIISEQQKDALMMKAFEYQLHGDDKMTLQVIHQSELMAYIKEIY 304

Qy        322 --------NLSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLP 373
                    |  || |: ||          :|   | : :|     |||  :::: :| :
Db        305 DMKKIPYLNPMELSNVINM-------FFEKV--IFNKDKPMGKESFL--RSVQEKFLHIQ 353

Qy        374 E----LLREMMSEQNGDALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQ 419
              :     | :| | | | : : ||     |  ||     ||: |    : |:
Db        354 KRSKILQQEEMDESNAEGVETI----QLKSLDDSTELEVNLPDFNSKDPE 399


RESULT 6
REGA_DICDI
ID   REGA_DICDI     STANDARD;      PRT;   793 AA.
AC   Q23917;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   3',5'-cyclic-nucleotide phosphodiesterase regA (EC 3.1.4.17) (PDEase
DE   regA).
GN   REGA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE=97140317; PubMed=8986798;
RA   Shaulsky G., Escalante R., Loomis W.F.;
RT   "Developmental signal transduction pathways uncovered by genetic
RT   suppressors.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:15260-15265(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX2;
RA   Thomason P.A., Traynor D., Cavet G., Chang W.T., Harwood A.J.,
RA   Kay R.R.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Morphological suppressor of tagB.
CC   -!- CATALYTIC ACTIVITY: Nucleoside 3',5'-cyclic phosphate + H(2)O =
CC       nucleoside 5'-phosphate.
CC   -!- DEVELOPMENTAL STAGE: Expressed at low levels in vegetative cells
CC       and at high levels in prespore and prestalk cells during
CC       development.
CC   -!- SIMILARITY: Belongs to the cyclic nucleotide phosphodiesterase
CC       family.
CC   -!- SIMILARITY: Contains 1 response regulatory domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U60170; AAB03508.1; -.
DR   EMBL; AJ005398; CAA06513.1; -.
DR   DictyBase; DDB0001583; regA.
DR   InterPro; IPR003607; Met_phsphohydro.
DR   InterPro; IPR002073; PDEase.
DR   InterPro; IPR001789; Response_reg.
DR   Pfam; PF00233; PDEase; 1.
DR   Pfam; PF00072; response_reg; 1.
DR   PRINTS; PR00387; PDIESTERASE1.
DR   ProDom; PD000039; Response_reg; 1.
DR   SMART; SM00471; HDc; 1.
DR   SMART; SM00448; REC; 1.
DR   PROSITE; PS00126; PDEASE_I; 1.
DR   PROSITE; PS50110; RESPONSE_REGULATORY; 1.
KW   Hydrolase; cAMP; cGMP.
FT   DOMAIN      161    280       RESPONSE REGULATORY.
FT   DOMAIN       18     28       POLY-SER.
FT   DOMAIN       52     69       POLY-ASN.
FT   DOMAIN       87     96       POLY-THR.
FT   DOMAIN      100    121       POLY-ASN.
FT   DOMAIN      166    169       POLY-ASP.
FT   DOMAIN      770    776       POLY-SER.
SQ   SEQUENCE   793 AA;  91175 MW;  6E065A96620F8C27 CRC64;

  Query Match             5.3%;  Score 125;  DB 1;  Length 793;
  Best Local Similarity   19.0%;  Pred. No. 2.2;
  Matches  107;  Conservative   88;  Mismatches  183;  Indels  184;  Gaps   27;

Qy          1 MPLSKHQIEQLSKPLS----------DDSICGVYLKLE----KSAFRPLRNEF------- 39
              | | :|:||:|:| :|          :  : |  :|    :|::    :|
Db        310 MELKEHEIEELTKKVSKMSSISKEAMESPLVSVTRNIEELLKQSSWSHYESEIKEKLSSI 369

Qy         40 -------NVAQTALRKLSQNPSADERDALQEACLNKWKILSDSLYEQFSKTTRDIELISW 92
                    |: : :  || :| | |              :    |  :|| ||
Db        370 LKELGSSNIYRPSFEKLIKNDSVD------------PVTKSFLVSEFSSTT-------- 408

Qy         93 FVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDD---KGKEREQADA 149
                      |  ||:               |  |  |    |      || |
Db        409 ------------SRRNSI--------------PTFPQTTYNRDTKEVIKGWE------ 434

Qy        150 KVKAFFQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAE-RKGEISQLKSMLTTTVAQE 208
                  |  :   ||:          :||: :: | :||  |  |  | :|:  :  |  |
Db        435 -----FDVFKYSEDD--------LMPLLVDM-FENFQLPEIFKIPIEKLQRFIMTVNALY 480

Qy        209 R-------FAIQFKME----------NAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAK 251
              |       |  |  :          || : :| ||  : |:|  || |     |  |
Db        481 RKNNRYHNFTHAFDVTQTVYTFLTSFNAAQYLTHLDIFALLISCMCHDLNHPGFNNTFQV 540

Qy        252 SLLTR----------VENALVHLSGIKLAPKAEAKTVE-------QEVAESSVS---EGE 291
              :  |            :||    |:  |:   :|   :|       :|:  | |      :
Db        541 NAQTELSLEYNDISVLENHHAMLT-FKILRNSECNILEGLNEDQYKELRRSVVQLILATD 599

Qy        292 LPSHMD-----TKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRDLAFHL-----L 341
              : :|  :        |:  :|   :: :   : :     | : |:::|: |   :||     |
Db        600 MQNHFEHTNKFQHHLNNLPF--DRNKKEDRQMILNFLIKCGDISNIAR--PWHLNFEWSL 655

Qy        342 REVSDYFRQSEPHS----PISFLLEKA------IRWGYL---------SLPELLREMM-- 380
              |   ::|:||   :    |::  ::|       |   ::         |: : |:|
Db        656 RVSDEFFQQSHYETICGYPVTPFMDKTKTTRARIAADFIDFVASPLFQSMAKFLKESQFL 715

Qy        381 ------SEQNGDALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPRSVE 434
                   : :|  |   :       |  |   |  :  ||: ::| : |:     :  :
Db        716 LKVISKNRENWQAYMELQKEGKCNDDD---LQFMEDPTILVKS-KLPKIDEEENRDKVSS 771

Qy        435 EHVSQTSPVDTQSKQDQKPQSS 456
                |:|: : |  : :  ||
Db        772 SSSSSTAPLTSTSSSNNETSSS 793


RESULT 7
SMC3_SCHPO
ID   SMC3_SCHPO     STANDARD;      PRT;  1194 AA.
AC   O42649;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Structural maintenance of chromosome 3 (Cohesin complex Psm3 subunit).
GN   PSM3 OR SMC3 OR SPAC10F6.09C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
```

```
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [2]
RP   FUNCTION, AND SUBCELLULAR LOCATION AND IDENTIFICATION IN A COHESIN
RP   COMPLEX WITH PSM1 AND RAD21.
RX   MEDLINE=20523814; PubMed=11069892;
RA   Tomonaga T., Nagao K., Kawasaki Y., Furuya K., Murakami A.,
RA   Morishita J., Yuasa T., Sutani T., Kearsey S.E., Uhlmann F.,
RA   Nasmyth K., Yanagida M.;
RT   "Characterization of fission yeast cohesin: essential anaphase
RT   proteolysis of Rad21 phosphorylated in the S phase.";
RL   Genes Dev. 14:2757-2770(2000).
CC   -!- FUNCTION: Involved in chromosome cohesion during cell cycle and in
CC       DNA repair. Central component of cohesin complex. The cohesin
CC       complex is required for the cohesion of sister chromatids after
CC       DNA replication. The cohesin complex apparently forms a large
CC       proteinaceous ring within which sister chromatids can be trapped.
CC       At anaphase, the complex is cleaved and dissociates from
CC       chromatin, allowing sister chromatids to segregate.
CC   -!- SUBUNIT: Cohesin complexes are composed of the psm1/smc1 and
CC       psm3/smc3 heterodimer attached via their hinge domain, rad21/scc1
CC       which link them, and psc3/scc3, which interacts with rad21.
CC   -!- SUBCELLULAR LOCATION: Nuclear protein. Associates with chromatin.
CC       Before prophase it is scattered along chromosome arms. At
CC       anaphase, the rad21 subunit of the cohesin complex is cleaved,
CC       leading to the dissociation of the complex from chromosomes,
CC       allowing chromosome separation.
CC   -!- DOMAIN: The flexible hinge domain, which separates the large
CC       intramolecular coiled coil regions, allows the heterotypic
CC       interaction with the corresponding domain of psm1, forming a V-
CC       shaped heterodimer. The two heads of the heterodimer are then
CC       connected by different ends of the cleavable rad21 protein,
CC       forming a ring structure (By similarity).
CC   -!- SIMILARITY: Belongs to the SMC family. SMC3 subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL009197; CAA15722.1; -.
DR   PIR; T37503; T37503.
DR   GeneDB_SPombe; SPAC10F6.09c; -.
DR   GO; GO:0005694; C:chromosome; IDA.
DR   GO; GO:0008278; C:cohesin complex; IDA.
DR   GO; GO:0007064; P:mitotic sister chromatid cohesion; IMP.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR003405; SMC_C.
DR   InterPro; IPR003395; SMC_N.
DR   Pfam; PF02483; SMC_C; 1.
DR   Pfam; PF02463; SMC_N; 1.
KW   Mitosis; Cell cycle; Chromosome partition; ATP-binding; Coiled coil;
KW   Nuclear protein.
FT   NP_BIND      32     39       ATP (POTENTIAL).
FT   DOMAIN      176    497       COILED COIL (POTENTIAL).
FT   DOMAIN      498    667       FLEXIBLE HINGE.
FT   DOMAIN      668    908       COILED COIL (POTENTIAL).
FT   DOMAIN      970   1006       COILED COIL (POTENTIAL).
FT   DOMAIN     1097   1132       ALA/ASP-RICH (DA-BOX).
SQ   SEQUENCE   1194 AA;  136850 MW;  FD4D1E81A9E2E423 CRC64;

  Query Match             5.3%;  Score 125;  DB 1;  Length 1194;
  Best Local Similarity   21.2%;  Pred. No. 3.9;
  Matches  113;  Conservative   84;  Mismatches  184;  Indels  152;  Gaps   26;

Qy          8 IEQLSKPLSDDSICGVYLKLEKS--------AFRPLRN--------------EFNVAQTA 45
              ||  |:      : |: |  ::|           :|  ||              :|:  | :
Db        627 IETASQYARSHQLNGITLSGDRSDKKGALTAGYRDYRNSRLDAIKNVKTYQIKFSDLQES 686

Qy         46 LRKL-SQNPSADER-----DALQEACLNKWKILSD---------------SLYEQFSKT 83
              |  |:  | |::      | ||:| |:  :   |                 | |
Db        687 LEKCRSEIESFDQKITACLDDLQKAQLSLKQFERDHIPLKDELVTITGETTDLQESMHHK 746

Qy         84 TRDIELISWFVAAQFLLDT-TLESAANSLEWLADLSEKHWDHLNPVLPVETLKS------ 136
              :| :||:          :|:  |||  || |:  ::|| :  | |:|   || |||
Db        747 SRMLELV--------VLELHTLEQQANDLK--SELSSE-MDELDP-KDVEALKSLSGQIE 794

Qy        137 -----DDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAERK 191
                  |  |||   :|:  |      : | :: ||  : : ||  |:        :: :
Db        795 NLSHEFDAIIKERAHIEARKTAL-----EYELNTNLY--LRRNPLKAEI------GSDNR 841

Qy        192 GEISQLKSMLTTTVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAK 251
              : |:| |:  : :  |   :|    ::      |: |::: :| | : | |
Db        842 IDESELNSVKRSLLKYEN-KLQIIKSSSSGLEEQMQRINSEISDKRNELESLEE---LQH 897

Qy        252 SLLTRVENALVHLSGIKLAPKAEAKTVEQEVAESSV-------------SEGELPSHMDT 298
              : ||:|                :||  |:  |: |:              | | ||
Db        898 EVATRIEQ-------------DAKINERNAAKRSLLLARKKECNEKIKSLGVLPEEAFI 943

Qy        299 KHIERIPMASEQAQTVSQHLHAGN--LSELGNLN----------NMNRDLAFHLLREVSD 346
              |::       |  :  : : ||  |  | : |::|              ||    ||   :
Db        944 KYV------STSSNAIVKKLHKINEALKDYGSVNKKAYEQFNNFTKQRD---SLLARREE 994

Qy        347 YFRQSEPHSPISFLL--------EKAIRWGYLSLPELLREMMSEQNGDALSTIFNAAGLN 398
              |  |  | :: :|         |:  :    |  |:  :::    |:    : |   :
Db        995 LRRSQESISELTTVLDQRKDEAIERTFKQVAKSFSEIFVKLVPAGRGE---LVMNRR--S 1049

Qy        399 HLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPRSVEEHVSQTSPVDTQSKQDQ 451
              | | :  ::|          ||  | |: :   :    ||  |  | |   :|
Db       1050 ELSQSIEQDISMDI------DTPSQKSSIDNYTGISIRVSFNSKDDEQLNINQ 1096


RESULT 8
MYH3_HUMAN
ID   MYH3_HUMAN     STANDARD;      PRT;  1940 AA.
AC   P11055; Q15492;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Myosin heavy chain, fast skeletal muscle, embryonic (Muscle embryonic
DE   myosin heavy chain) (SMHCE).
GN   MYH3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89263803; PubMed=2726495;
```

```
RA   Eller M.S., Stedman H.H., Sylvester J.E., Fertels S.H.,
RA   Rubinstein N.A., Kelly A.M., Sarkar S.;
RT   "Nucleotide sequence of full length human embryonic myosin heavy
RT   chain cDNA.";
RL   Nucleic Acids Res. 17:3591-3592(1989).
RN   [2]
RP   SEQUENCE OF 774-1940 FROM N.A.
RX   MEDLINE=90033298; PubMed=2806546;
RA   Eller M.S., Stedman H.H., Sylvester J.E., Fertels S.H., Wu Q.-L.,
RA   Raychowdhury M.K., Rubinstein N.A., Kelly A.M., Sarkar S.;
RT   "Human embryonic myosin heavy chain cDNA. Interspecies sequence
RT   conservation of the myosin rod, chromosomal locus and isoform
RT   specific transcription of the gene.";
RL   FEBS Lett. 256:21-28(1989).
RN   [3]
RP   SEQUENCE OF 856-1940 FROM N.A.
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=90235862; PubMed=1691980;
RA   Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
RA   Arnold H.H.;
RT   "Identification of three developmentally controlled isoforms of human
RT   myosin heavy chains.";
RL   Eur. J. Biochem. 189:55-65(1990).
RN   [4]
RP   SEQUENCE OF 856-1940 FROM N.A.
RX   MEDLINE=89366648; PubMed=2771643;
RA   Karsch-Mizrachi I., Travis M., Blau H., Leinwand L.A.;
RT   "Expression and DNA sequence analysis of a human embryonic skeletal
RT   muscle myosin heavy chain gene.";
RL   Nucleic Acids Res. 17:6167-6179(1989).
CC   -!- FUNCTION: Muscle contraction.
CC   -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
CC       heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
CC       and 2 regulatory light chain subunits (MLC-2).
CC   -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC   -!- DEVELOPMENTAL STAGE: Abundantly present in fetal skeletal muscle
CC       and not present or barely detectable in heart and adult skeletal
CC       muscle.
CC   -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC       cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC       characteristic for alpha-helical coiled coils.
CC   -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
CC       meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be
CC       split further into 2 globular subfragments (S1) and 1 rod-shaped
CC       subfragment (S2).
CC   -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC   -!- SIMILARITY: Contains 1 IQ domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13988; CAA32167.1; -.
DR   EMBL; X13100; CAA31492.1; -.
DR   EMBL; X51593; CAA35942.1; -.
DR   EMBL; X15696; CAA33731.1; -.
DR   PIR; S04090; S04090.
DR   HSSP; P13538; 2MYS.
DR   Genew; HGNC:7573; MYH3.
DR   MIM; 160720; -.
DR   GO; GO:0007517; P:muscle development; TAS.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW   Calmodulin-binding; ATP-binding; Methylation; Multigene family.
FT   DOMAIN        1    781       MYOSIN HEAD-LIKE.
FT   DOMAIN      782    811       IQ.
FT   DOMAIN      840   1933       COILED COIL (POTENTIAL).
FT   NP_BIND     179    186       ATP (POTENTIAL).
FT   DOMAIN      656    678       ACTIN-BINDING.
FT   DOMAIN      758    772       ACTIN-BINDING.
FT   MOD_RES     130    130       METHYLATION (TRI-) (POTENTIAL).
FT   CONFLICT   1331   1331       A -> G (IN REF. 3).
FT   CONFLICT   1391   1392       KK -> QE (IN REF. 1 AND 2).
FT   CONFLICT   1608   1609       SR -> RA (IN REF. 3).
FT   CONFLICT   1663   1664       RG -> QT (IN REF. 2).
SQ   SEQUENCE   1940 AA;  224035 MW;  43CA58C6A4BA1253 CRC64;

  Query Match             5.3%;  Score 124.5;  DB 1;  Length 1940;
  Best Local Similarity   20.6%;  Pred. No. 8;
  Matches  100;  Conservative   86;  Mismatches  200;  Indels   99;  Gaps   20;

Qy          7 QIEQLSKPLSDDSICGVYLKLEKSAFR----PLRNEFNVAQTALRKLSQNPSADERDALQ 62
              :|:::::   |:       |  :|          |: : :  :  | |:  :   | |
Db        919 KIKEVTERAEDEEEINAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATE----- 973

Qy         63 EACLNKWKILSDSLY---EQFSKTTRDIELISWFVAAQFLLD--TTLESAANSLEWLADL 117
                 || | |:: |     |  :| ||: : :    | |  ||     |   |||
Db        974 ----NKVKNLTEELSGLDETIAKLTREKKALQ--EAHQQALDDLQAEEDKVNSLNKTKSK 1027

Qy        118 SEKHWDHLNPVLPVE-TLKSDDDKGKEREQADAKV--KAFFQLVGDSEESSILYAPVLQL 174
              |:  : |   | .|  |: | :: | : : | |:  ::   |  | ::
Db       1028 LEQQVEDLESSLEQEKKLRVDLERNKRKLEGDLKLAQESILDLENDKQQ---------- 1076

Qy        175 PLVGEVTFFDFQSAERKGEISQLKSMLTTTVAQERFAIQF--KMENAKRCVTQL-DRLSA 231
                    | :  ::  |  ||:| :     ::   :||  |::  :  : :| : : |
Db       1077 --------LDERLKKKDFEYCQLQSKVED---EQTLGLQFQKKIKELQARIEELEEEIEA 1125

Qy        232 LVSTKCHSLGSQSTNFGFAKSLLTRVENA-LVHLSGIKLAPKAEAK--TVEQEVAESSVS 288
              :|:  :   :|      : |  |:| |  |  : |:|  | ||:   : ::: |:::
Db       1126 ERATRAKTEKQRSDYARELEELSERLEEAGGVTSTQIELNKKREAEFLKLRRDLEEATLQ 1185

Qy        289 EGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRDL-----AFHLLRE 343
              : : :  || : :   ||                 : ||  : : |      | | |
Db       1186 HEAMVATLRKKHADSVAELGEQ--------------IDNLQRVKQKLEKEKSEFKL--E 1228

Qy        344 VSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDALSTIFNAAGLNHLDQV 403
              : |     |  |    |||  |         | : :||           | | |   |
Db       1229 IDDLSSSMESVSKSKANLEKICR--------TLEDQLSE-----------ARGKNEEIQR 1269

Qy        404 LLPEVST------PTVGIESPQTPQAKPSVSD-PRSVEEHVSQTSPVDTQSKQDQKPQSS 456
              | |::|        |  | |  : :  ||   ||  :    ||  :  | ::: | :::
Db       1270 SLSELTTQKSRLQTEAGELSRQLEEKESIVSQLSRSKQAFTQQTEELKRQLEEENKAKNA 1329

Qy        457 ATSAL 461
                ||
Db       1330 LAHAL 1334


RESULT 9
MYH3_RAT
ID   MYH3_RAT       STANDARD;      PRT;  1940 AA.
AC   P12847;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Myosin heavy chain, fast skeletal muscle, embryonic.
GN   MYH3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
```

```
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87060988; PubMed=3783701;
RA   Strehler E.E., Strehler-Page M.-A., Perriard J.C., Periasamy M.,
RA   Nadal-Ginard B.;
RT   "Complete nucleotide and encoded amino acid sequence of a mammalian
RT   myosin heavy chain gene. Evidence against intron-dependent evolution
RT   of the rod.";
RL   J. Mol. Biol. 190:291-317(1986).
CC   -!- FUNCTION: Muscle contraction.
CC   -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
CC       heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
CC       and 2 regulatory light chain subunits (MLC-2).
CC   -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC   -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC       cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC       characteristic for alpha-helical coiled coils.
CC   -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
CC       meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be
CC       split further into 2 globular subfragments (S1) and 1 rod-shaped
CC       subfragment (S2).
CC   -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC   -!- SIMILARITY: Contains 1 IQ domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04267; CAA27817.1; -.
DR   PIR; A24922; A24922.
DR   HSSP; P13538; 2MYS.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW   Calmodulin-binding; ATP-binding; Methylation; Multigene family.
FT   DOMAIN        1    781       MYOSIN HEAD-LIKE.
FT   DOMAIN      782    811       IQ.
FT   DOMAIN      840   1933       COILED COIL (POTENTIAL).
FT   NP_BIND     179    186       ATP (POTENTIAL).
FT   DOMAIN      656    678       ACTIN-BINDING.
FT   DOMAIN      758    772       ACTIN-BINDING.
FT   MOD_RES     130    130       METHYLATION (TRI-) (POTENTIAL).
SQ   SEQUENCE   1940 AA;  223857 MW;  B5D546A596E5A696 CRC64;

  Query Match             5.3%;  Score 124.5;  DB 1;  Length 1940;
  Best Local Similarity   20.9%;  Pred. No. 8;
  Matches  101;  Conservative   90;  Mismatches  194;  Indels   99;  Gaps   21;

Qy          7 QIEQLSKPLSDDSICGVYLKLEKSAFR----PLRNEFNVAQTALRKLSQNPSADERDALQ 62
              :|:::::   |:       |  :|       |: : :  :  | |: :   | |
Db        919 KIKEVTERAEDEEEINAELTAKKRKLEDECSELKKDIDDLELTLAKVEKEKHATE----- 973

Qy         63 EACLNKWKILSDSLY---EQFSKTTRDIELISWFVAAQFLLD--TTLESAANSLEWLADL 117
                 || | |:: |    |  :| ||: : :    | |  ||     |   |||  |
Db        974 ----NKVKNLTEELAGLDETIAKLTREKKALQ--EAHQQTLDDLQAEEDKVNSLSKLKSK 1027

Qy        118 SEKHWDHLNPVLPVE-TLKSDDDKGKEREQADAKV--KAFFQLVGDSEESSILYAPVLQL 174
              |:  | |   |  |  |: | :: | : : | |:  ::   |  | ::
Db       1028 LEQQVDDLESSLEQEKKLRVDLERNKRKLEGDLKLAQESILDLENDKQQ---------- 1076

Qy        175 PLVGEVTFFDFQSAERKGEISQLKSMLTTTVAQER---FAIQFKMENAKRCVTQL-DRLS 230
                  | :  ::  | |||:|     |  |:     :| |::  :  : :| : :
Db       1077 --------LDERLKKKDFEYSQLQS----KVEDEQTLSLQLQKKIKELQARIEELEEEIE 1124

Qy        231 ALVSTKCHSLGSQSTNFGFAKSLLTRVENA-LVHLSGIKLAPKAEAK--TVEQEVAESSV 287
              |  :|:  :   :|      : |  |:| |  |  : |:|  | ||:   : ::: |:::
Db       1125 AERATRAKTEKQRSDYARELEELSERLEEAGGVTSTQIELNKKREAEFLKLRRDLEEATL 1184

Qy        288 SEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRDL-----AFHLLR 342
                  : :  || :    |:| |:             :: ||  : : |       | |
Db       1185 QHEATVATLRKKHAD---SAAELAE-----------QIDNLQRVKQKLEKEKSEFKL-- 1227

Qy        343 EVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDALSTIFNAAGLNHLDQ 402
              |: |      | |      |||  |          | : :||           | | |   |
Db       1228 EIDDLSSSVESVSKSKANLEKICR--------TLEDQLSE----------ARGKNEETQ 1268

Qy        403 VLLPEVSTPTVGIESPQTPQAKPSVSDPRSVEEHVSQTSPVDTQSKQDQKPQ----SSAT 458
                | |::|    ::: :  :    : :  |:   :|::     ||  :: | |    : |
Db       1269 RSLSELTTQKSRLQT-EAGELSRQLEEKESIVSQLSRSKQAFTQQIEELKRQLEEENKAK 1327

Qy        459 SALS 462
              :||:
Db       1328 NALA 1331


RESULT 10
MYSP_SCHMA
ID   MYSP_SCHMA     STANDARD;      PRT;   866 AA.
AC   P06198;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Paramyosin.
OS   Schistosoma mansoni (Blood fluke).
OC   Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC   Schistosomatoidea; Schistosomatidae; Schistosoma.
OX   NCBI_TaxID=6183;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91270282; PubMed=2052029;
RA   Laclette J.P., Landa A., Arcos L., Willms K., Davis A.E.,
RA   Shoemaker C.B.;
RT   "Paramyosin is the Schistosoma mansoni (Trematoda) homologue of
RT   antigen B from Taenia solium (Cestoda).";
RL   Mol. Biochem. Parasitol. 44:287-296(1991).
RN   [2]
RP   SEQUENCE OF 303-742 FROM N.A.
RX   MEDLINE=87018840; PubMed=3094144;
RA   Lanar D.E., Pearce E.J., James S.L., Sher A.;
RT   "Identification of paramyosin as schistosome antigen recognized by
RT   intradermally vaccinated mice.";
RL   Science 234:593-596(1986).
CC   -!- FUNCTION: Paramyosin is a major structural component of many thick
CC       filaments isolated from invertebrate muscles.
CC   -!- SUBUNIT: Homodimer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC   -!- SIMILARITY: HIGH, TO MYOSIN HEAVY CHAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M35499; AAA29915.1; -.
DR   EMBL; M14163; AAA29914.1; -.
DR   InterPro; IPR002928; Myosin_tail.
```

```
DR   Pfam; PF01576; Myosin_tail; 1.
KW   Coiled coil; Muscle protein; Thick filament; Myosin.
FT   DOMAIN        1     22       NONHELICAL REGION (POTENTIAL).
FT   DOMAIN       23    839       COILED COIL (POTENTIAL).
FT   DOMAIN      840    866       NONHELICAL REGION (POTENTIAL).
FT   DISULFID    750    750       INTERCHAIN (POTENTIAL).
FT   CONFLICT    423    428       DQVKEL -> GSSQRI (IN REF. 2).
FT   CONFLICT    430    431       SS -> KL (IN REF. 2).
FT   CONFLICT    637    637       V -> L (IN REF. 2).
FT   CONFLICT    639    639       T -> I (IN REF. 2).
FT   CONFLICT    691    691       G -> E (IN REF. 2).
FT   CONFLICT    720    720       S -> F (IN REF. 2).
SQ   SEQUENCE   866 AA;  100387 MW;  42FA5E6E78176AE0 CRC64;

  Query Match             5.3%;  Score 124;  DB 1;  Length 866;
  Best Local Similarity   21.3%;  Pred. No. 2.9;
  Matches  103;  Conservative   82;  Mismatches  182;  Indels  116;  Gaps   21;

Qy          7 QIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACL 66
              ::| | |                | ||     :    ||       |    ||: :    :: |
Db        122 EVENLQKQKG---------KAEKDKSHLIMEVDNVLGQLDGALKAKQSAESKLEGLDSQL 172

Qy         67 NKWKILSDSLYEQFSK--------TTRDIEL--ISWFVAAQFLLDTTLESAANSLEWLAD 116
               |: | |:| |  |.::       |: : ||  |:     || |       | :|||   |
Db        173 NRLKSLTDDLQRQLTELNNAKSRLTSENFELLHINQDYEAQIL---NYSKAKSSLESQVD 229

Qy        117 LSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGD---------SEESSIL 167
                          :  :| ||: | |     |:: :  |:  |          |||:| |
Db        230 --------------DLKRSLDDEAKNRFNLQAQLTS-LQMDYDNLQAKYDEESEEASNL 273

Qy        168 YAPVLQLPLVGEVTFFDFQSAERKGEISQLKSMLTTTVAQERFAIQFKMENAKRCVTQLD 227
                :          :|: |:   |  | :    : ::: |   |    :| |       :|:|:
Db        274 RS---------QVSKFNADIAALKSKFE--RELMSKTEEFEEMKRKFTMR-----ITELE 317

Qy        228 RLSALVSTKCHSLGSQSTNFGF-AKSLLTRVENALVHLSGIKLAPKAEAKTVEQEVAESS 286
                :     | ||    |       | | : :|:  : |   :|  :|:|       |||
Db        318 DTAERERLKAVSLEKLKTKLTLEIKDLQSEIES--LSLENSELIRRAKA-------AESL 368

Qy        287 VSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNL-SELGNLNN-MNRDLAFHLLREV 344
                |  :|   :|   ||   : |: :|  |::|   :| ::| : || : |:      |::
Db        369 AS--DLQRRVDELTIEVNTLTSQNSQLESENLRLKSLVNDLTDKNNLLERE-----NRQM 421

Qy        345 SDYFRQSEPHSPISFLLEKAIRWGYLSLPEL--LREMMSEQNGDALSTIFNA-------- 394
                :|  ::          |: ::|     | :|  ||  :  :  :  | : :|
Db        422 NDQVKE----------LKSSLRDANRRLTDLEALRSQLEAERDNLASALHDAEEALHDMD 471

Qy        395 -------AGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPRSVEEHVSQTSPVDTQS 447
                      | ||||   :   :      :|| :        |  |::||     : :: :
Db        472 QKYQASQAALNHLKSEMEQRLRERDEELESLRK-------STTRTIEELTVTITEMEVKY 524

Qy        448 KQD 450
                | :
Db        525 KSE 527


RESULT 11
DPO1_RICPR
ID   DPO1_RICPR     STANDARD;      PRT;   867 AA.
AC   O05949;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   DNA polymerase I (EC 2.7.7.7) (POL I).
GN   POLA OR RP776.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=782;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Madrid E;
RX   MEDLINE=97419517; PubMed=9274032;
RA   Andersson J.O., Andersson S.G.E.;
RT   "Genomic rearrangements during evolution of the obligate
RT   intracellular parasite Rickettsia prowazekii as inferred from an
RT   analysis of 52015 bp nucleotide sequence.";
RL   Microbiology 143:2783-2795(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Madrid E;
RX   MEDLINE=99039499; PubMed=9823893;
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Madrid E, and B;
RX   MEDLINE=99416441; PubMed=10486973;
RA   Andersson J.O., Andersson S.G.E.;
RT   "Genome degradation is an ongoing process in Rickettsia.";
RL   Mol. Biol. Evol. 16:1178-1191(1999).
CC   -!- FUNCTION: In addition to polymerase activity, This DNA polymerase
CC       exhibits 3' to 5' and 5' to 3' exonuclease activity (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC       + {DNA}(N).
CC   -!- SUBUNIT: Single-chain monomer with multiple functions (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the DNA polymerase type-A family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Y11784; CAA72476.1; -.
DR   EMBL; AJ235273; CAA15203.1; -.
DR   EMBL; AJ238755; CAB56085.1; -.
DR   EMBL; AJ238756; CAB56089.1; -.
DR   PIR; C71638; C71638.
DR   HSSP; P00582; 1KFS.
DR   InterPro; IPR008918; 5_3_exo_C.
DR   InterPro; IPR002421; 5_3_exonuclease.
DR   InterPro; IPR001098; DNA_pol.
DR   InterPro; IPR002298; DNA_polI.
DR   InterPro; IPR000513; Exo_N_I.
DR   Pfam; PF02739; 5_3_exonuc_N; 1.
DR   Pfam; PF01367; 5_3_exonuclease; 1.
DR   Pfam; PF00476; DNA_pol_A; 1.
DR   PRINTS; PR00868; DNAPOLI.
DR   SMART; SM00475; 53EXOc; 1.
DR   SMART; SM00279; HhH2; 1.
DR   SMART; SM00482; POLAc; 1.
DR   TIGRFAMs; TIGR00593; pola; 1.
DR   PROSITE; PS00447; DNA_POLYMERASE_A; 1.
KW   Transferase; DNA-directed DNA polymerase; DNA replication; DNA repair;
KW   Hydrolase; Exonuclease; DNA-binding; Complete proteome.
FT   DOMAIN        1    283       EXONUCLEASE.
SQ   SEQUENCE   867 AA;  98651 MW;  557194B38FB69BA2 CRC64;

  Query Match             5.3%;  Score 123.5;  DB 1;  Length 867;
  Best Local Similarity   21.2%;  Pred. No. 3.1;
  Matches   90;  Conservative   60;  Mismatches  172;  Indels  103;  Gaps   17;

Qy         18 DSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACLNKWKILSDSLY 77
              |:| ||     |:|   | :|   :     | |  |  :|  || |
Db        186 DNIPGVPSIGPKTA-SSLITKFGSVENIFNSLDQISSIKQRKTLQNA------------ 231

Qy         78 EQFSKTTRDIELISWFVAAQFLLDTT--LESAANSLEWLADLSEKHWDHLNPVLPVETLK 135
                     |:  |||| :     ||:   |:   |:|:|    |:|    |   |    |
```

```
Db        232 -------REAALISWKLIG---LDSNVDLDFDLNNLKWSPPNSKK----LTGFLQEYGFK 277

Qy        136 SDDDKGKEREQADAKVKAFFQL-VGDSEESSILYAPVLQLPLVGEVTFFDFQSAERKGEI 194

Db        278 S----------LYKRVENLFDIKINDHEE-----------------IVDNKVTEAK-EI 308

Qy        195 SQLKSMLTTTVAQER------FAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFG 248

Db        309 SNASELANFAKEAERIGIFGIYLLQQKGEN-RALILSLQNQSYIIKITNNNYNIKNNNDW 367

Qy        249 FAKSLLTRVEN--------ALVHLSGIKLAPKAEAKTVEQ-EVAESSVSEGELPSHMDTK 299

Db        368 FSHIILNLLTNKSIKKITYSLKHLLKFYANQSHQITAIEDLELMQYALSAGLVQKNLFTK 427

Qy        300 HIERIPMASEQAQTVSQ--HLHAGNLSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPI 357

Db        428 TLTKDNIINESARIVINFISLYKQTLLEL------QKNKAFRLYREID---------LPT 472

Qy        358 SFLLEKAIRWGYLSLPELLREMMSEQNGDAL----------STIFNAAGLNHLDQVLLPE 407

Db        473 CFILDKMEKVGIKVDANYLNRLSDEFGTEILKIEEEIFALSGTKFNIGSQKQLGEILFKK 532

Qy        408 VSTPT 412

Db        533 MQLPS 537
```

```
RESULT 12
MYH7_HUMAN
ID   MYH7_HUMAN     STANDARD;      PRT;  1935 AA.
AC   P12883; Q14904; Q16579;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Myosin heavy chain, cardiac muscle beta isoform (MyHC-beta).
GN   MYH7 OR MYHCB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91065634; PubMed=2249844;
RA   Jaenicke T., Diederich K.W., Haas W., Schleich J., Lichter P.,
RA   Pfordt M., Bach A., Vosberg H.P.;
RT   "The complete sequence of the human beta-myosin heavy chain gene and
RT   a comparative analysis of its product.";
RL   Genomics 8:194-206(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90301496; PubMed=2362820;
RA   Liew C.-C., Sole M.J., Yamauchi-Takihara K., Kellam B.,
RA   Anderson D.H., Lin L., Liew J.;
RT   "Complete sequence and organization of the human cardiac beta-myosin
RT   heavy chain gene.";
RL   Nucleic Acids Res. 18:3647-3651(1990).
RN   [3]
RP   SEQUENCE OF 1-115 FROM N.A.
RX   MEDLINE=89264452; PubMed=2726733;
RA   Yamauchi-Takihara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RT   "Characterization of human cardiac myosin heavy chain genes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).
RN   [4]
RP   ERRATUM.
RA   Yamauchi-Takihara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RL   Proc. Natl. Acad. Sci. U.S.A. 86:7416-7417(1989).
RN   [5]
RP   SEQUENCE OF 1310-1935 FROM N.A.
RX   MEDLINE=86176778; PubMed=2421254;
RA   Saez L., Leinwand L.A.;
RT   "Characterization of diverse forms of myosin heavy chain expressed in
RT   adult human skeletal muscle.";
RL   Nucleic Acids Res. 14:2951-2969(1986).
RN   [6]
RP   REVISIONS.
RA   Leinwand L.A.;
RL   Submitted (MAR-1988) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE OF 1410-1935 FROM N.A.
RX   MEDLINE=88299163; PubMed=2969919;
RA   Kurabayashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.;
RT   "Molecular cloning and characterization of human cardiac alpha- and
RT   beta-form myosin heavy chain complementary DNA clones. Regulation of
RT   expression during development and pressure overload in human
RT   atrium.";
RL   J. Clin. Invest. 82:524-531(1988).
RN   [8]
RP   SEQUENCE OF 785-1935 FROM N.A.
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=90235862; PubMed=1691980;
RA   Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
RA   Arnold H.H.;
RT   "Identification of three developmentally controlled isoforms of human
RT   myosin heavy chains.";
RL   Eur. J. Biochem. 189:55-65(1990).
RN   [9]
RP   SEQUENCE OF 1393-1935 FROM N.A.
RX   MEDLINE=87192738; PubMed=3032769;
RA   Jandreski M.A., Liew C.-C.;
RT   "Construction of a human ventricular cDNA library and
RT   characterization of a beta myosin heavy chain cDNA clone.";
RL   Hum. Genet. 76:47-53(1987).
RN   [10]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=96039076; PubMed=8533830;
RA   Arai S., Matsuoka R., Hirayama K., Sukurai H., Tamura M., Ozawa T.,
RA   Kimura M., Imamura S.-I., Furutani Y., Joh-O K., Kawana M., Takao A.,
RA   Hosoda S., Momma K.;
RT   "Missense mutation of the beta-cardiac myosin heavy-chain gene in
RT   hypertrophic cardiomyopathy.";
RL   Am. J. Med. Genet. 58:267-276(1995).
RN   [11]
RP   VARIANTS CMH1 GLU-256 AND ARG-741.
RX   MEDLINE=93248216; PubMed=8483915;
RA   Fananapazir L., Dalakas M.C., Cyran F., Cohn G., Epstein N.D.;
RT   "Missense mutations in the beta-myosin heavy-chain gene cause central
RT   core disease in hypertrophic cardiomyopathy.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:3993-3997(1993).
RN   [12]
RP   VARIANT CMH1 GLN-403.
RX   MEDLINE=90367131; PubMed=1975517;
RA   Geisterfer-Lowrance A.A.T., Kass S., Tanigawa G., Vosberg H.-P.,
RA   McKenna W., Seidman C.E., Seidman J.G.;
RT   "A molecular basis for familial hypertrophic cardiomyopathy: a beta
RT   cardiac myosin heavy chain gene missense mutation.";
RL   Cell 62:999-1006(1990).
RN   [13]
RP   VARIANTS CMH1 GLN-249; GLN-403; CYS-453 AND MET-606.
RX   MEDLINE=92204193; PubMed=1552912;
RA   Watkins H., Rosenzweig A., Hwang D.S., Levi T., McKenna W.,
RA   Seidmann C.E., Seidmann J.G.;
RT   "Characteristics and prognostic implications of myosin missense
RT   mutations in familial hypertrophic cardiomyopathy.";
RL   New Engl. J. Med. 326:1108-1114(1992).
RN   [14]
RP   VARIANTS CMH1 GLN-403; CYS-453; ARG-584 AND MET-606.
RX   MEDLINE=94070863; PubMed=8250038;
RA   Watkins H., Thierfelder L., Anan R., Jarcho J., Matsumori A.,
RA   McKenna W., Seidman J.G., Seidman C.E.;
RT   "Independent origin of identical beta cardiac myosin heavy-chain
RT   mutations in hypertrophic cardiomyopathy.";
RL   Am. J. Hum. Genet. 53:1180-1185(1993).
RN   [15]
RP   VARIANTS CMH1 GLN-403 AND VAL-908.
RX   MEDLINE=92346810; PubMed=1638703;
RA   Epstein N.D., Cohn G.M., Cyran F., Fananapazir L.;
```

```
RT   "Differences in clinical expression of hypertrophic cardiomyopathy
RT   associated with two distinct mutations in the beta-myosin heavy chain
RT   gene. A 908Leu-->Val mutation and a 403Arg-->Gln mutation.";
RL   Circulation 86:345-352(1992).
RN   [16]
RP   VARIANTS CMH1 LEU-403 AND TRP-403.
RX   MEDLINE=94075629; PubMed=8254035;
RA   Dausse E., Komajda M., Fetler L., Dubourg O., Dufour C., Carrier L.,
RA   Wisnewsky C., Bercovici J., Hengstenberg C., Al-Mahdawi S.;
RT   "Familial hypertrophic cardiomyopathy. Microsatellite haplotyping and
RT   identification of a hot spot for mutations in the beta-myosin heavy
RT   chain gene.";
RL   J. Clin. Invest. 92:2807-2813(1993).
RN   [17]
RP   VARIANT CMH1 TRP-403.
RX   MEDLINE=94093568; PubMed=8268932;
RA   Moolman J.C., Brink P.A., Corfield V.A.;
RT   "Identification of a new missense mutation at Arg403, a CpG mutation
RT   hotspot, in exon 13 of the beta-myosin heavy chain gene in
RT   hypertrophic cardiomyopathy.";
RL   Hum. Mol. Genet. 2:1731-1732(1993).
RN   [18]
RP   VARIANT CMH1 ASN-615.
RX   MEDLINE=93038688; PubMed=1417858;
RA   Nishi H., Kimura A., Harada H., Toshima H., Sasazuki T.;
RT   "Novel missense mutation in cardiac beta myosin heavy chain gene
RT   found in a Japanese patient with hypertrophic cardiomyopathy.";
RL   Biochem. Biophys. Res. Commun. 188:379-387(1992).
RN   [19]
RP   VARIANT CMH1 GLY-778.
RX   MEDLINE=93343938; PubMed=8343162;
RA   Harada H., Kimura A., Nishi H., Sasazuki T., Toshima H.;
RT   "A missense mutation of cardiac beta-myosin heavy chain gene linked
RT   to familial hypertrophic cardiomyopathy in affected Japanese
RT   families.";
RL   Biochem. Biophys. Res. Commun. 194:791-798(1993).
RN   [20]
RP   VARIANT CMH1 VAL-908.
RX   MEDLINE=93168485; PubMed=8435239;
RA   Al-Mahdawi S., Chamberlain S., Cleland J., Nihoyannopoulos P.,
RA   Gilligan D., French J., Choudhury L., Williamson R., Oakley C.;
RT   "Identification of a mutation in the beta cardiac myosin heavy chain
RT   gene in a family with hypertrophic cardiomyopathy.";
RL   Br. Heart J. 69:136-141(1993).
RN   [21]
RP   VARIANT CMH1 TRP-719.
RX   MEDLINE=95179132; PubMed=7874131;
RA   Greve G., Bachinski L., Friedman D.L., Czernuzewicz G., Anan R.,
RA   Towbin J., Seidman C.E., Roberts R.;
RT   "Isolation of a de novo mutant myocardial beta MHC protein in a
RT   pedigree with hypertrophic cardiomyopathy.";
RL   Hum. Mol. Genet. 3:2073-2075(1994).
RN   [22]
RP   VARIANTS CMH1 CYS-513; ARG-716 AND TRP-719.
RX   MEDLINE=94110336; PubMed=8282798;
RA   Anan R., Greve G., Thierfelder L., Watkins H., McKenna W., Solomon S.,
RA   Vecchio C., Shono H., Nakao S., Tanaka H., Mares A. Jr., Towbin J.A.,
RA   Spirito P., Roberts R., Seidman J.G., Seidman C.E.;
RT   "Prognostic implications of novel beta cardiac myosin heavy chain gene
RT   mutations that cause familial hypertrophic cardiomyopathy.";
RL   J. Clin. Invest. 93:280-285(1994).
RN   [23]
RP   VARIANT CMH1 THR-797.
RX   MEDLINE=96047159; PubMed=7581410;
RA   Moolman J.C., Brink P.A., Corfield V.A.;
RT   "Identification of a novel Ala797Thr mutation in exon 21 of the beta-
RT   myosin heavy chain gene in hypertrophic cardiomyopathy.";
RL   Hum. Mutat. 6:197-198(1995).
RN   [24]
RP   VARIANT CMH1 CYS-453.
RX   MEDLINE=96209901; PubMed=8655135;
RA   Ko Y.-L., Chen J.-J., Tang T.-K., Cheng J.-J., Lin S.-Y., Liou Y.-C.,
RA   Kuan P., Wu C.-W., Lien W.-P., Liew C.-C.;
RT   "Malignant familial hypertrophic cardiomyopathy in a family with a
RT   453Arg-->Cys mutation in the beta-myosin heavy chain gene:
RT   coexistence of sudden death and end-stage heart failure.";
RL   Hum. Genet. 97:585-590(1996).
RN   [25]
RP   VARIANTS CMH1 THR-349 AND TRP-719.
RX   MEDLINE=98204402; PubMed=9544842;
RA   Jeschke B., Uhl K., Weist B., Schroder D., Meitinger T.,
RA   Dohlemann C., Vosberg H.-P.;
RT   "A high risk phenotype of hypertrophic cardiomyopathy associated with

  Query Match             5.2%;  Score 123;  DB 1;  Length 1935;
  Best Local Similarity   22.1%;  Pred. No. 9.8;
  Matches   94;  Conservative   63;  Mismatches  140;  Indels  128;  Gaps   19;

Qy          7 QIEQLSKPL------SDDSICGVYLK---LEKSAFRPLRNEF----------NVAQTALR 47
              ::|:  | |       :::::  |  |   |||:  | |:||           | |  ||
Db       1385 ELEEAKKKLAQRLQEAEEAVEAVNAKCSSLEKTKHR-LQNEIEDLMVDVERSNAAAAALD 1443

Qy         48 KLSQN------------PSADERDALQEAC--------------------LNKWKILSD 74
              |  :|              |  | :: |:                      |  :|  :
Db       1444 KKQRNFDKILAEWKQKYEESQSELESSQKEARSLSTELFKLKNAYEESLEHLETFKRENK 1503

Qy         75 SLYEQFS----------KTTRDIELISWFVAAQFL-LDTTLESAANSLEWLADLSEKHWD 123
              :| |: |           ||  ::| :   : |: : | : || |  |||      :
Db       1504 NLQEEISDLTEQLGSSGKTIHELEKVRKQLEAEKMELQSALEEAEASLE------HEEGK 1557

Qy        124 HLNPVLPVETLKSD-----DDKGKEREQA---DAKVKAFFQLVGDSEESSILYAPVLQLP 175
              |   |    :|::       :| :| |||     :|     |   |:|  |   |  ::
Db       1558 ILRAQLEFNQIKAEIERKLAEKDEEMEQAKRNHLRVVDSLQTSLDAETRSRNEALRVKKK 1617

Qy        176 LVGEVTFFDFQ-------SAERKGEISQLKSMLTTTVAQ------------ERFAI---- 212
              : |::   : |        :|| : ::  |:|:|  |  |             |  ||
Db       1618 MEGDLNEMEIQLSHANRMAAEAQKQVKSLQSLLKDTQIQLDDAVRANDDLKENIAIVERR 1677

Qy        213 ----QFKMENAKRCVTQLDRLSAL-------VSTKCHSLGSQSTNFGFAK--------SL 253
                  | ::|  :  | | :|   | .       | :   | ||:|:    |         |
Db       1678 NNLLQAELEELRAVVEQTERSRKLAEQELIETSERVQLLHSQNTSLINQKKKMDADLSQL 1737

Qy        254 LTRVENALVHLSGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTK-HIERIPMASEQAQ 312
              | || |       ::    || |  :: : ::::   ||    ||  |:||:    ||  .
Db       1738 QTEVEEA------VQECRNAEEK-AKKAITDAAMMAEELKKEQDTSAHLERMKKNMEQTI 1790

Qy        313 TVSQH 317
                ||
Db       1791 KDLQH 1795


RESULT 13
MYH7_RAT
ID   MYH7_RAT       STANDARD;      PRT;  1935 AA.
AC   P02564;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Myosin heavy chain, cardiac muscle beta isoform (MyHC-beta).
GN   MYH7.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=90016823; PubMed=2798112;
RA   Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
RT   "Complete nucleotide sequence of full length cDNA for rat beta
RT   cardiac myosin heavy chain.";
RL   Nucleic Acids Res. 17:7529-7530(1989).
RN   [2]
RP   DISCUSSION OF SEQUENCE.
RX   MEDLINE=90133919; PubMed=2614840;
RA   McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
```

```
RT   "Full-length rat alpha and beta cardiac myosin heavy chain sequences.
RT   Comparisons suggest a molecular basis for functional differences.";
RL   J. Mol. Biol. 210:665-671(1989).
RN   [3]
RP   SEQUENCE OF 1524-1935 FROM N.A.
RX   MEDLINE=82220036; PubMed=7045682;
RA   Mahdavi V., Periasamy M., Nadal-Ginard B.;
RT   "Molecular characterization of two myosin heavy chain genes expressed
RT   in the adult heart.";
RL   Nature 297:659-664(1982).
RN   [4]
RP   SEQUENCE OF 1871-1935 FROM N.A.
RC   STRAIN=Wistar; TISSUE=Heart;
RX   MEDLINE=85179510; PubMed=6241892;
RA   Mahdavi V., Lompre A.M., Chambers A.P., Nadal-Ginard B.;
RT   "Cardiac myosin heavy chain isozymic transitions during development
RT   and under pathological conditions are regulated at the level of mRNA
RT   availability.";
RL   Eur. Heart J. 5:181-191(1984).
CC   -!- FUNCTION: Muscle contraction.
CC   -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
CC       heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
CC       and 2 regulatory light chain subunits (MLC-2).
CC   -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC   -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC       cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC       characteristic for alpha-helical coiled coils.
CC   -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
CC       meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be
CC       split further into 2 globular subfragments (S1) and 1 rod-shaped
CC       subfragment (S2).
CC   -!- MISCELLANEOUS: The cardiac alpha isoform is a 'fast' ATPase
CC       myosin, while the beta isoform is a 'slow' ATPase.
CC   -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC   -!- SIMILARITY: Contains 1 IQ domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X15939; CAA34065.1; -.
DR   EMBL; J00752; AAA41654.1; -.
DR   EMBL; M32698; AAA41659.1; -.
DR   PIR; S06006; S06006.
DR   HSSP; P08799; 1MND.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR004009; Myosin_N.
DR   InterPro; IPR002928; Myosin_tail.
DR   Pfam; PF00612; IQ; 2.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   Pfam; PF01576; Myosin_tail; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 1.
KW   Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW   ATP-binding; Methylation; Multigene family; Calmodulin-binding.
FT   DOMAIN        1    780       MYOSIN HEAD-LIKE.
FT   DOMAIN      781    810       IQ.
FT   DOMAIN      840   1935       COILED COIL (POTENTIAL).
FT   NP_BIND     178    185       ATP.
FT   DOMAIN      655    677       ACTIN-BINDING.
FT   DOMAIN      757    771       ACTIN-BINDING.
FT   MOD_RES     129    129       METHYLATION (TRI-) (POTENTIAL).
FT   CONFLICT   1529   1531       IRK -> VRR (IN REF. 3).
FT   CONFLICT   1731   1731       D -> H (IN REF. 3).
FT   CONFLICT   1784   1784       N -> K (IN REF. 3).
FT   CONFLICT   1851   1851       T -> N (IN REF. 3).
FT   CONFLICT   1858   1858       R -> K (IN REF. 3).
SQ   SEQUENCE   1935 AA;  223082 MW;  C8376C324A7BD82B CRC64;

  Query Match             5.2%;  Score 123;  DB 1;  Length 1935;
  Best Local Similarity   22.1%;  Pred. No. 9.8;
  Matches  102;  Conservative   65;  Mismatches  154;  Indels  140;  Gaps   21;

Qy          7 QIEQLSKPL------SDDSICGVYLK---LEKSAFRPLRNEF----------NVAQTALR 47
              ::|:  | |      :::::  |  |   |||:  | |:||           | |  ||
Db       1385 ELEEAKKKLAQRLQDAEEAVEAVNAKCSSLEKTKHR-LQNEIEDLMVDVERSNAAAAALD 1443

Qy         48 KLSQN------------PSADERDALQEAC--------------------LNKWKILSD 74
              |  :|             |  | :: |:                     |  :|  :
Db       1444 KKQRNFDKILVEWKQKYEESQSELESSQKEARSLSTELFKLKNAYEESLEHLETFKRENK 1503

Qy         75 SLYEQFS----------KTTRDIELISWFVAAQFL-LDTTLESAANSLEWLADLSEKHWD 123
              :| |: |          |:  ::| |   : |: | | : || |  |||      :
Db       1504 NLQEEISDLTEQLGSTGKSIHELEKIRKQLEAEKLELQSALEEAEASLE------HEEGK 1557

Qy        124 HLNPVLPVETLKSD-----DDKGKEREQA---DAKVKAFFQLVGDSEESSILYAPVLQLP 175
              |   |    :|::      :| :| |||     :|    |   |:|  |   |  ::
Db       1558 ILRAQLEFNQIKAEIERKLAEKDEEMEQAKRNHLRVVDSLQTSLDAETRSRNEALRVKKK 1617

Qy        176 LVGEVTFFDFQ-------SAERKGEISQLKSMLTTTVAQ------------ERFAI---- 212
              : |::   : |        :|| : ::  |:|:|  |  |            |  ||
Db       1618 MEGDLNEMEIQLSHANRMAAEAQKQVKSLQSLLKDTQIQLDDAVRANDDLKENIAIVERR 1677

Qy        213 ----QFKMENAKRCVTQLDRLSAL-------VSTKCHSLGSQSTNFGFAK--------SL 253
                 | ::|  :  | | :|  |          | :   | ||:|:|:    |          |
Db       1678 NNLLQAELEELRAVVEQTERSRKLAEQELIETSERVQLLHSQNTSLINQKKKMDADLSQL 1737

Qy        254 LTRVENALVHLSGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTK-HIERIPMASEQAQ 312
              | || |      ::    || |  :: : ::::   ||    ||  |:||  | :  |
Db       1738 QTEVEEA------VQECRNAEEK-AKKAITDAAMMAEELKKEQDTSAHLER--MKNNMEQ 1788

Qy        313 TVSQHLH----------AGNLSELGNLNNMNRDLAFHLLRE 343
              |:    |          |   :|  |    |:|   | |
Db       1789 TIKDLQHRLDEAEQIALKGGKKQLQKLEARVRELENELEAE 1829


RESULT 14
ANT1_ONCVO
ID   ANT1_ONCVO     STANDARD;      PRT;  2022 AA.
AC   P21249;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Major antigen.
GN   OVT1.
OS   Onchocerca volvulus.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC   Onchocercidae; Onchocerca.
OX   NCBI_TaxID=6282;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95287898; PubMed=7770081;
RA   Triteeraprapab S., Richie T.L., Tuan R.S., Shepley K.J., Dinman J.D.,
RA   Neubert T.A., Scott A.L.;
RT   "Molecular cloning of a gene expressed during early embryonic
RT   development in Onchocerca volvulus.";
RL   Mol. Biochem. Parasitol. 69:161-171(1995).
RN   [2]
RP   SEQUENCE OF 733-866 FROM N.A.
RX   MEDLINE=89127417; PubMed=2464764;
RA   Donelson J.E., Duke B.O.L., Moser D., Zeng W., Erondu N.E.,
RA   Lucius R., Renz A., Karam M., Flores G.Z.;
RT   "Construction of Onchocerca volvulus cDNA libraries and partial
RT   characterization of the cDNA for a major antigen.";
RL   Mol. Biochem. Parasitol. 31:241-250(1988).
CC   -!- FUNCTION: May be a myofibrillar protein.
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U12681; AAA80009.1; -.
DR   EMBL; J03995; AAA29412.1; -.
DR   PIR; T43214; T43214.
KW   Antigen; Coiled coil.
FT   DOMAIN       74    120       COILED COIL (POTENTIAL).
FT   DOMAIN      151    251       COILED COIL (POTENTIAL).
FT   DOMAIN      327    384       COILED COIL (POTENTIAL).
FT   DOMAIN      417   1879       COILED COIL (POTENTIAL).
SQ   SEQUENCE   2022 AA;  237341 MW;  B7132AACF1520317 CRC64;

  Query Match             5.2%;  Score 123;  DB 1;  Length 2022;
  Best Local Similarity   22.9%;  Pred. No. 10;
  Matches   85;  Conservative   64;  Mismatches  140;  Indels   82;  Gaps   16;

Qy         35 LRNEFNVAQTALRKLSQNPSADERDALQEACLNKWKILSDSLYEQFSKTTRDIELISWFV 94
              ||:: | |: |:  | ||     ||:: |   | ||  ||:|  :  :  || ||:|
Db        971 LRSKLNDAERAMADL-QN-----RDSILERENNDWKEKSDALNMELDR-LRD-ELLSVRR 1022

Qy         95 AAQFLL---DTTLESAA-NSLEWLADLSEKHWDHLNPVLPVETLKSDDDKGKEREQADAK 150
               |:  :   :| |::|| | :: |   : :    ||                      |: |
Db       1023 DAEKEINRYNTDLQTAARNEIKLLTPTNNEMKSQLN------------------AAEDK 1063

Qy        151 VKAFFQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAERKGEISQLKSMLTTTVAQERF 210
              : :  ::: | :             | |||   : :  : || :: |:| | ||    |
Db       1064 INSLNKVITDQQNKI--------RDLTGEVHHLEGELKDAKGNVANLESELDTT----RE 1111

Qy        211 AIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVHLSGIKLA.270
              |   |      |:|:::         |   : ||          || :  |   | :
Db       1112 RIHLLGEQNASLQTELNKIK----------GDIDSLFG---------ENDM--LKTAKES 1150

Qy        271 PKAEAKTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLN 330
              :||   ::|::  |  :  :    :|    |  :          |:|:   : :  ||
Db       1151 NEAEIDRLKQKLQRSIENAKKYSDALDKLRPEYDRL---------QNLYREKIKQAENLT 1201

Qy        331 NMNRDLAFHL------LREVSDYFRQSE-PHSPISFLLEKAIRWGYLSLPELLREMMSEQ 383
                :||   |       ||: :|    ||   : :   :||          :|||:  :::
Db       1202 QAVQDLESRLNQSRRELRDATDKLIASEGDRNALRSEVEKLQHEVQFMREQLLRK--TDE 1259

Qy        384 NGDALSTIFNA 394
                ||| : ||
Db       1260 YQAALSDLVNA 1270


RESULT 15
LBN_MOUSE
ID   LBN_MOUSE      STANDARD;      PRT;  1220 AA.
AC   Q8K1G2; Q8BRF3;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Limbin.
GN   EVC2 OR LBN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC   TISSUE=Limb bud;
RX   MEDLINE=22155879; PubMed=12136126;
RA   Takeda H., Takami M., Oguni T., Tsuji T., Yoneda K., Sato H.,
RA   Ihara N., Itoh T., Kata S.R., Mishina Y., Womack J.E., Moritomo Y.,
RA   Sugimoto Y., Kunieda T.;
RT   "Positional cloning of the gene limbin responsible for bovine
RT   chondrodysplastic dwarfism.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:10549-10554(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Kidney;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 21-744 FROM N.A.
RC   STRAIN=C57BL/6J;
RX   MEDLINE=22354683; PubMed=12466851;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
CC   -!- FUNCTION: Plays a critical role in bone formation and skeletal
CC       development. May be involved in early embryonic morphogenesis.
CC   -!- TISSUE SPECIFICITY: Expressed in long and cranial bones, kidney
CC       and heart. Strongly expressed in proliferating chondrocytes,
CC       osteoblasts and osteoclasts.
CC   -!- DEVELOPMENTAL STAGE: Found in the embryo at day E7, E11, E15, and
CC       E17. At the limb bud formation stage E11, it is expressed in fore-
CC       and hindlimb buds, branchial arches, and facial primordia.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; AB083066; BAC06589.1; -.
DR   EMBL; BC037473; AAH37473.1; -.
DR   EMBL; AK044977; BAC32167.1; ALT_INIT.
DR   MGD; MGI:1915775; Evc2.
KW   Transmembrane; Coiled coil.
FT   TRANSMEM     211    231       POTENTIAL.
FT   DOMAIN       355    404       COILED COIL (POTENTIAL).
FT   DOMAIN       563    644       COILED COIL (POTENTIAL).
FT   DOMAIN       854    875       COILED COIL (POTENTIAL).
FT   DOMAIN       920   1005       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1220 AA;  137638 MW;  E67671714A866B1D CRC64;

  Query Match             5.2%;  Score 122.5;  DB 1;  Length 1220;
  Best Local Similarity  19.2%;  Pred. No. 5.7;
  Matches  113;  Conservative   76;  Mismatches  198;  Indels  201;  Gaps   23;

Qy          3 LSKHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQ--NPSADERDA 60
              | ||: ||  : :|                    |    : |: |:: | |  :  |:   |
Db        590 LQKHK-EQQKEQVS-----------------LGEASSTAEDAVQYLHQWRSVMAEHTAA 630

Qy         61 LQE-------ACLNKWKILSDSLYEQFSKTTRDI-------ELIS----WFVAAQFLLDT 102
              |:|       | |:  ::|: || |: ::  | :       ||:     |    | | :
Db        631 LEELQERLDQAALDDLRVLTVSLSEKATEELRRLQSTAMTQELLKRSAPWLFLQQILEEH 690

Qy        103 TLESAANSLEWLADLSE---------------------------KHWDHL--NPVLPVE 132
              : |||| :  :  |:  |                          :||:||    :
Db        691 SRESAARTTQLEAEERERGQELVQGVRQRLQQDALEAYTEEQAELRHWEHLVFMKLCCAA 750

Qy        133 TLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLV-GEVTFFDFQSAERK 191
                |::|  : |::|      | ||    : |       | || |    |      | | |:
Db        751 ISLSEEDLLRVRQEAQG---CFSQL----DRS-------LALPRVRARVLQQQAQMAWRE 796

Qy        192 GEISQLKSMLTTTVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAK 251
              |  :|   |     | :                        : |  | |     | |
Db        797 AEFRKLDQALAAPELQSK----------------------ARKLRSKGR-----GKAD 827

Qy        252 SLLTRVENAL---VHLSGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTKHIERIPMAS 308
              |   :|: :      : ::|| :   : ::: |       |      :    :::  |:
Db        828 LLKKNLEDKIRLFEERAPVELADQVRGELLQERVQRLEAQEAHFAESLVALQFQKVARAA 887

Qy        309 EQAQTVS-----QHLHAGNLS-----------------ELGNLNNMNRDLAFHLL--- 341
              |      :     | |  | ||                 ||  |  : | |   |:
Db        888 ETLSVYTALLSIQDLLLGELSESETLTKSACVQILESHRPELQELQELERKLEDQLVQQE 947

Qy        342 -----REVSDYFR-----------QSEPHSPISFLLEKAIRWGYLSLPELLREMMSE-Q 383
                  | :  : |           : :|   :| :| :|:  |   | :  : :  | |
Db        948 EAEQQRVLESWQRWAADGPGLSEPEEMDPERQVSAILRQALNKGQKLLEQHQQRVREEWQ 1007

Qy        384 NGDALS--------------------------TIFNAAGLNHLDQVLLPEVSTPTVGI 415
              ||  |                          |:   : |  |  |:|| |
Db       1008 NGAVLEDSLESIEADTMASLCSQGLRLVSYLSRMTMVPGSTLLRLLSVVLPAAS------ 1061

Qy        416 ESPQTPQAKPSVSDPRSVEEHVSQTSPVDTQSKQDQKPQSSATSALSW 463
               ||      :||     |:|   |:  ::  :| |  ||       |
Db       1062 -QPQLLALLDAVS-----EKHSDHTAENESSGEQAQAEQSKRRKHQVW 1103


Search completed: July  3, 2004, 02:44:55
Job time : 22 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 3, 2004, 02:06:04 ; Search time 72 Seconds
(without alignments)
2028.957 Million cell updates/sec

Title: US-10-780-347-2
Perfect score: 2343
Sequence: 1 MPLSKHQIEQLSKPLSDDSI.....DTQSKQDQKPQSSATSALSW 463

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	291	12.4	377	16	Q8D6T7	Q8d6t7 vibrio vuln
2	284.5	12.1	373	16	Q87HC9	Q87hc9 vibrio para
3	252	10.8	358	2	Q93ED3	Q93ed3 rhizobium l
4	241.5	10.3	351	16	Q8U7V6	Q8u7v6 agrobacteri
5	232.5	9.9	337	16	Q8XRS2	Q8xrs2 ralstonia s
6	209	8.9	366	16	Q9I1B9	Q9i1b9 pseudomonas
7	174	7.4	361	16	Q88IB1	Q88ib1 pseudomonas
8	152.5	6.5	1454	4	Q9UPS3	Q9ups3 homo sapien
9	152.5	6.5	1503	4	Q8IWU2	Q8iwu2 homo sapien
10	151	6.4	2245	5	Q86A36	Q86a36 dictyosteli
11	143.5	6.1	959	5	Q86IW2	Q86iw2 dictyosteli
12	143	6.1	884	5	O02239	O02239 caenorhabdi
13	141.5	6.0	344	16	Q9I750	Q9i750 pseudomonas
14	141	6.0	1534	5	Q8MPV7	Q8mpv7 caenorhabdi
15	141	6.0	1538	5	Q8MPV6	Q8mpv6 caenorhabdi
16	140.5	6.0	3742	8	Q9TKW8	Q9tkw8 nephroselmi
17	137	5.8	347	16	Q8PF73	Q8pf73 xanthomonas
18	136.5	5.8	7735	5	Q9VMT8	Q9vmt8 drosophila
19	136	5.8	1216	16	Q8EAA0	Q8eaa0 shewanella
20	134.5	5.7	1179	16	Q8DPJ9	Q8dpj9 streptococc
21	133.5	5.7	357	16	Q7UL73	Q7ul73 rhodopirell
22	130	5.5	16223	5	Q8IR22	Q8ir22 drosophila
23	129.5	5.5	630	5	Q25613	Q25613 onchocerca
24	129.5	5.5	1133	5	Q21022	Q21022 caenorhabdi
25	127.5	5.4	2471	12	Q9YTK3	Q9ytk3 ateline her
26	127	5.4	1935	6	Q9BE39	Q9be39 bos taurus
27	126.5	5.4	752	4	Q13597	Q13597 homo sapien
28	126.5	5.4	893	4	Q8N6Z0	Q8n6z0 homo sapien
29	126.5	5.4	1023	16	Q92B64	Q92b64 listeria in
30	126.5	5.4	2094	11	Q80Y35	Q80y35 mus musculu
31	126	5.4	613	2	Q9XD53	Q9xd53 moraxella c
32	126	5.4	1029	16	Q81DR0	Q81dr0 bacillus ce
33	125.5	5.4	1992	13	Q04834	Q04834 xenopus lae
34	125.5	5.4	2167	2	Q845L8	Q84518 mycoplasma
35	125	5.3	1935	6	Q9GKR1	Q9gkr1 sus scrofa
36	124.5	5.3	550	3	O59725	O59725 schizosacch
37	124.5	5.3	1964	13	O93522	O93522 xenopus lae
38	124	5.3	1935	6	Q8MJU9	Q8mju9 equus cabal
39	123.5	5.3	1051	2	Q49524	Q49524 mycoplasma
40	123	5.2	611	4	Q14905	Q14905 homo sapien
41	123	5.2	800	10	Q94GW1	Q94gw1 oryza sativ
42	123	5.2	1935	4	Q9H1D5	Q9h1d5 homo sapien
43	123	5.2	2471	13	Q9DDN8	Q9ddn8 xenopus lae
44	122.5	5.2	1220	11	Q8K1G2	Q8k1g2 mus musculu
45	122	5.2	1344	2	Q49545	Q49545 mycoplasma

ALIGNMENTS

RESULT 1
Q8D6T7
ID Q8D6T7 PRELIMINARY; PRT; 377 AA.
AC Q8D6T7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Uncharacterized protein conserved in bacteria.
GN VV20438.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AE016809; AAO07392.1; -.
KW Complete proteome.
SQ SEQUENCE 377 AA; 42931 MW; A04DA3EA65C4679E CRC64;

Query Match 12.4%; Score 291; DB 16; Length 377;
Best Local Similarity 21.8%; Pred. No. 3.5e-12;
Matches 91; Conservative 84; Mismatches 161; Indels 82; Gaps 14;

```
Qy          1 MPLSKHQIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDA 60
              :|:|  |   :  | || :   :| :|:   :             | : :|   :|
Db         16 LPISAEQPAGVD-PRSDITPQSLYYRLKDQRMQ------------ARNIERNAIIEEEPI 62

Qy         61 LQEACLNKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEK 120
              |  | | |::  | : :  :  |:|:| ::| : |  ||  :  |  |  |     |
Db         63 LNYA--NLWQVFLDEVPQALTSQTKDLEYVAWLIEALTRLD-GFKGMAVGYELATYYLEN 119

Qy        121 HWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLVGEV 180
              :|  | |:          |: | |       ::      | |   | :::: |:  :|:   :
Db        120 YWADLYPM--------PDEDGLE-----TRISPIIGLNGIDNEGTLIF-PLSCIPMTEGL 165
```

```
Qy        181 -----TFFDFQS-------------AERKGEISQLKSMLTTTVAQERFAIQFKMENAKRC 222
Db        166 LEQPYAYWEYQQALDLERCDEDKKRAKRDQGVPELSDITQAVKSSSSQFYQRLIEDLNRA 225

Qy        223 VTQLDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVHLSGIKLAPKAEAKTVEQEV 282
Db        226 ISSFERFSQALDSACEEV----TPSSYISRKLDAIHSAIMHLASDKM-----------I 269

Qy        283 AESSVSEGELPSHMDTKH---IERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRDLAFH 339
Db        270 KVSKETNSDEPSHSASSHEGEIDGRGQGEANFQLLAANMH------------SREQAIV 316

Qy        340 LLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDALSTIFNAAGL 397
Db        317 HLQRVADFFRETEPHSPVSYTIEQVIRWCGMSLPELLAELISD--GDAKKGYFRLVGI 372

RESULT 2
Q87HC9
ID   Q87HC9       PRELIMINARY;      PRT;   373 AA.
AC   Q87HC9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein.
GN   VPA1036.
OS   Vibrio parahaemolyticus.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=670;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIMD 2210633 / Serotype O3:K6;
RX   MEDLINE=22508454; PubMed=12620739;
RA   Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA   Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA   Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT   "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT   distinct from that of V. cholerae.";
RL   Lancet 361:743-749(2003).
DR   EMBL; AP005087; BAC62379.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   373 AA;  42347 MW;  07D369BC6C3E59E3 CRC64;

  Query Match             12.1%;  Score 284.5;  DB 16;  Length 373;
  Best Local Similarity   24.3%;  Pred. No. 9.8e-12;
  Matches  105;  Conservative   81;  Mismatches  129;  Indels  117;  Gaps   19;

Qy          9 EQLSKPLSDDSICGVYLKLE---KSAFRPLRNEFNVAQTALRKLSQNPSADERDA-LQEA 64
Db         10 EQVLAPISEDNPTGIDPREDVSPQSAYYRLKDQRMVARNA-----------ERNAIIEEE 58

Qy         65 CL----NKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEK 120
Db         59 SIHTHNNLWRVFVEEVPEVLATQSKDFEFVAWLIEALTRL-YGFRGMGVGYKIASAFVES 117

Qy        121 HWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLVG-- 178
Db        118 YWDELHPM--------PDEDGIE-----TRISAITGLNGIDSEGTLIF-PLASIPLTDMG 163

Qy        179 ---EVTFFDFQSA---------ERKG-------EISQLKSMLTTT--------VAQERFA 211
Db        164 VEQAYAYWEYQQAIELERLDEDKRRARVDQGAIELSRITDTVKSTSDEFYQSLIADLEFA 223

Qy        212 ------IQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVHLS 265
Db        224 TGAFNEFSQRLDDAVGDVTP----SSYISKRIES----------------IHAALKHLL 262

Qy        266 GIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSE 325
Db        263 GNRFKDKAEV----------SIS--------DENGDEELEEASQQAVT-------HNSSA 297

Qy        326 LGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNG 385
Db        298 LLVTNMQSREQAIQQLQHVADFFRESEPHSPVSYSIEQIIRWCGMPLPELLAELISD--G 355

Qy        386 DALSTIFNAAGL 397
Db        356 DAKRSYFRLVGI 367

RESULT 3
Q93ED3
ID   Q93ED3       PRELIMINARY;      PRT;   358 AA.
AC   Q93ED3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ImpA.
GN   IMPA.
OS   Rhizobium leguminosarum (biovar trifolii).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX   NCBI_TaxID=386;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bladergroen M.R., Badelt K., Stronk O.P., Lugtenberg E.J.J.,
RA   Spaink H.P.;
RT   "An avirulence locus of a symbiotic Rhizobium leguminosarum strain is
RT   involved in temperature-dependent protein secretion.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF361470; AAL17808.1; -.
SQ   SEQUENCE   358 AA;  39039 MW;  9F5C107C58D22395 CRC64;

  Query Match             10.8%;  Score 252;  DB 2;  Length 358;
  Best Local Similarity   23.9%;  Pred. No. 1.7e-09;
  Matches   96;  Conservative   71;  Mismatches  170;  Indels   64;  Gaps   14;

Qy          8 IEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACLN 67
Db          3 VREIIDPLEQNHPCGENVR-SNTAFREIYYRIKDARNAARTAERSIIPGETITIAPA--- 58

Qy         68 KWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLAD-------LSEK 120
Db         59 -WHDVSNLGLQLLSSKSKDIEVLAWLAEAQLRL--------RGFSGLHDVYVATVSLLDK 109

Qy        121 HWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLVG-- 178
Db        110 HFDALHSI-------GDGD-------VEERFAPFAGLNGVGGEGTLIQAIRLTSLIPGGK 155

Qy        179 --EVTFFDFQSAERKGEISQLKSMLTTTVAQERFAIQFKMENAKRCVTQLDRLSALVSTK 236
Db        156 FAQFSLWDFQLSQRPNETERRKELQQAAAEAGVAQMSAYLDVLTGCIAAFDRMVEILDEH 215

Qy        237 CHSLGSQSTNFGFAKSLLTRVENALVHLSGIKLAPKAEAKTVEQEVAESSVSEGELPSHM 296
Db        216 CGDQAPPSSN---TRNVLQEAASAIRMLAGIE-AP-TEAPLAAPQALASQGSE-PLPA-S 268

Qy        297 DTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSP 356
Db        269 ETDGSEAIGLRPVSAEMI-----------------RSREEAFELLIAVARYFRRTEPHSP 311

Qy        357 ISFLLEKAIRWGYLSLPELLREMMSEQNGDALSTIFNAAGL 397
Db        312 ISMSIETLVRRGRMDFFELLAELLPEQ--QTRNAVLTAAGI 350

RESULT 4
Q8U7V6
ID   Q8U7V6       PRELIMINARY;      PRT;   351 AA.
AC   Q8U7V6;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein Atu4343.
```

```
GN   ATU4343 OR AGR_L_1042.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608550; PubMed=11743193;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA   Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA   Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA   Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA   Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA   Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA   Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA   Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA   Nester E.W.;
RT   "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT   C58.";
RL   Science 294:2317-2323(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608551; PubMed=11743194;
RA   Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA   Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA   Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA   Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA   Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA   Cielo C., Slater S.;
RT   "Genome sequence of the plant pathogen and biotechnology agent
RT   Agrobacterium tumefaciens C58.";
RL   Science 294:2323-2328(2001).
DR   EMBL; AE009363; AAL45137.1; -.
DR   EMBL; AE008251; AAK89097.1; -.
DR   PIR; AC3090; AC3090.
DR   PIR; G98196; G98196.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   351 AA;  38490 MW;  85460F21E1C63FC7 CRC64;

  Query Match             10.3%;  Score 241.5;  DB 16;  Length 351;
  Best Local Similarity   23.6%;  Pred. No. 9e-09;
  Matches   96;  Conservative   64;  Mismatches  162;  Indels   85;  Gaps   15;

Qy          5 KHQIEQLSKPLSDDSICGVYLKLE---KSAFRPLRNEFNVAQTALRKLSQNPSADERDAL 61
              |  || |          ||  :: :   :  :  :::| | |:           |:|| |
Db          7 KRDIEFLGN-------CGDNIRNDSRTREIYYRIKDERNQAR----------AEERAAS 48

Qy         62 QEACL---NKWKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLS 118
               :  |   : |  :|:   :     ::|:|:::|   |   |           |   ||
Db         49 PQDNLKISSSWDSVSNLGLQIIYSESKDVEILAWLAEASLRL-RGFHGLREIYELCGDLF 107

Qy        119 EKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLV- 177
              ||| |  :       |||:       : |   | | |   | : |  |:    |:
Db        108 YNHWDSLRSI-------SDDND-------EEKFAPFAGLNGIGSEGT-LVQPLRLASLIP 152

Qy        178 ----GEVTFFDFQSAERKGEISQLKSMLTTTVAQERFAIQFKMENAKRCVTQLDRLSALV 233
                  || : :||| |:|  |  : : :       |:   :     |::  | ::|::
Db        153 GGKFGEHSLWDFQLAQRPNESKRREELYRIASEAGVAAMSSHLAAVNTCLSSFDAITAVL 212

Qy        234 STKCHSLGSQSTNFGFAKSLLTRVENALVHLSGIKLAPKAEAKTV---EQEVAESSVSEG 290
              | :|      |:|          : | |:  :       |  :|:   :|| |
Db        213 SERCGQAAPPSSN----------IRNTLIEAA-------AAIRTLGGRDQEPAP------ 249

Qy        291 ELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRDLAFHLLREVSDYFRQ 350
                      :|: |  :  | |       |   |         :|| ||  |  |: |||:
Db        250 ----------VEQTPAIA--AGTDESGQSAARTSPASPEGISSRDEAFETLLSVARYFRR 297

Qy        351 SEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDALSTIFNAAGL 397
              :|||||||  :|  :| | :   ||| |:: |    | : :  |||:
Db        298 TEPHSPISLSIETLVRRGRMDFSELLAELLPET--QARNAVLTAAGI 342
```

```
RESULT 5
Q8XRS2
ID   Q8XRS2      PRELIMINARY;      PRT;   337 AA.
AC   Q8XRS2;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein RSp0759.
GN   RSP0759 OR RS01949.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OG   Plasmid megaplasmid.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Burkholderiaceae; Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646080; CAD17910.1; -.
DR   GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW   Plasmid; Hypothetical protein; Complete proteome.
SQ   SEQUENCE   337 AA;  37094 MW;  C1EE68258BB39B97 CRC64;

  Query Match             9.9%;  Score 232.5;  DB 16;  Length 337;
  Best Local Similarity   24.9%;  Pred. No. 3.6e-08;
  Matches   99;  Conservative   60;  Mismatches  127;  Indels  111;  Gaps   18;

Qy         38 EFNVAQTALRKLSQNPSADERDALQEACLNKWKILSDSLYEQFSKTTRDIELISWFVAAQ 97
              ||:  | | |    :|| |: | : |     |: :         | |:|: | :|   |
Db          6 EFDTIQEARR--FDDPSLDQGDWVTEIKEADWRAVVSESTALLQKRTKDLRLAAWLTEA- 62

Qy         98 FLLDTTLESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQL 157
                 :          | :| | |::|:|| | ||           | :  :|:  :   |
Db         63 LSKERGFAGLREGYELIAGLCEQYWEHLFP-LP-----------EPDDPEARTGSMAWL 109

Qy        158 VGDSEESSILYAPVLQLPLV----GEVTFFDFQSA-----------ERKGEISQLKSMLT 202
               | | :       : ::|||    |   : |:: |           ::  |:|: |
Db        110 AGRSSQL------IREVPLVEAGNGGFNYVDWEVATNLIEAIRRDPDQADELSRGK---- 159

Qy        203 TTVAQERFAIQFKMENAKR----------------CVTQLDRLSALVSTKCHSLGSQSTN 246
               | |::|      | |:|                |   | || :::  :    |  : :
Db        160 --VTQDQF------EAARRASSSAFYTALYNDVVGCGAALLRLESVLDGRA---GDHAPS 208

Qy        247 FGFAKSLLTRVENALVHLSGIKL-----APKAEAKTVEQEVAESSVSEGELPSHMDTKHI 301
              |  |:  |  |  |||   | |       | : :| |  |: |:
Db        209 FRAAREALQAV-RALVERFGGKTETKPSAVEPKATTAPQQTAQ---------------- 250

Qy        302 ERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLL 361
                |:|     ||           :| :   |  | : |||:::|||:|||||:::|
Db        251 ---PVAEAVPDTV----------VGPIR--TRAQALNQLREVAEFFRQTEPHSPVAYLA 294

Qy        362 EKAIRWGYLSLPELLREMMSEQNGDA-LSTIFNAAGL 397
              :| :|| : |   || :: :   || || |    ||
Db        295 ARAAKWGDMPLHAWLRAVVKD---DATLSQIEELLGL 328
```

```
RESULT 6
Q9I1B9
ID   Q9I1B9      PRELIMINARY;      PRT;   366 AA.
AC   Q9I1B9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein PA2360.
GN   PA2360.
```

```
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004662; AAG05748.1; -.
DR   PIR; D83350; D83350.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   366 AA;  40816 MW;  E32CF1DB0118536A CRC64;

  Query Match             8.9%;  Score 209;  DB 16;  Length 366;
  Best Local Similarity   21.5%;  Pred. No. 1.8e-06;
  Matches   91;  Conservative   73;  Mismatches  151;  Indels  108;  Gaps   16;

Qy          1 MPLSKH--QIEQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADER 58
              :||| :   :| | :|:    ||  |: :   |           ||:| :   :
Db          3 LPLSGNALSLEVLLEPIDPGQACGPSLRYDPDYDR-----------LRELRREDDSSLP 50

Qy         59 DALQEACLNK--WKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLAD 116
               : :|   :  | :    |   : ::|: | :|   | :|     |     :| ||:
Db         51 TGVWQAEAKRADWAAVEQLASELLQRRSKDLMLAAWLGEA-WLQRGGLGGLQRALVLLAE 109

Qy        117 LSEKHWDHLNPVLPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPL 176
              | |:: : ::|         | |:        |  :: : :|:              :|||
Db        110 LCERYPEEVHP------QAQDGDQSWRVPPIDWLLRRYAELLH-----------TRLPL 151

Qy        177 VG-----EVTFFDFQSAERK----GEISQLKSMLTTTVAQERFAIQFKMENAKR------ 221
              :|     |:| : :|  :|:    |:    |: |     |::     |:: | |
Db        152 MGQGAFAEITLYAWQRLQRQQVASGDSKSAKAALEAAQLQQK-----KLDEALRAEPLVQ 206

Qy        222 ----------CVTQL-------DRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVHL 264
                       |  ||       ||    ::  |  |          |: |     |   :|
Db        207 WQRKQASLLACQQQLLRLEQWCDRCLGELAPSCQPLREV-----IAQWLALLKEFIAMH- 260

Qy        265 SGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLS 324
                |:|     :  |||:  |||       ||    | :| ::      :
Db        261 ------PQAPLSEEQPPVAEADASEG------DTDGEESVPASAPSGPAGAP-------- 300

Qy        325 ELGNLNNMNRDLAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQN 384
                  :|: |:  |  ::||  ::|||||: :|:::|: ||   | ||| |::   |
Db        301 -------TSREDAYRQLLLIADYLARTEPHSPVPYLIKRAVEWGNKPLSELLAELI---N 350

Qy        385 GDA 387
                |:
Db        351 ADS 353


RESULT 7
Q88IB1
ID   Q88IB1      PRELIMINARY;      PRT;   361 AA.
AC   Q88IB1;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Conserved hypothetical protein.
GN   PP3088.
OS   Pseudomonas putida (strain KT2440).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=160488;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22423060; PubMed=12534463;
RA   Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA   Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA   Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
RA   Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
RA   Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzez A.,
RA   Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
RA   Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA   Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuemmler B.,
RA   Fraser C.M.;
RT   "Complete genome sequence and comparative analysis of the
RT   metabolically versatile Pseudomonas putida KT2440.";
RL   Environ. Microbiol. 4:799-808(2002).
DR   EMBL; AE016785; AAN68696.1; -.
DR   TIGR; PP3088; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   361 AA;  39853 MW;  D822BFFD4F9C20BA CRC64;

  Query Match             7.4%;  Score 174;  DB 16;  Length 361;
  Best Local Similarity   20.6%;  Pred. No. 0.00048;
  Matches   74;  Conservative   58;  Mismatches  132;  Indels   96;  Gaps   11;

Qy         69 WKILSDSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNPV 128
              |: : :   :  :: ::|: |::| :           |     ||    |     |: | |
Db         59 WEQVINLSAQGLAQQSKDLMLVAW-LCEGLAQREHFEGITFGLELTEQLLVAFWETLYPT 117

Qy        129 LPVETLKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLVGEVTFFD---- 184
              |       :|   ||    : :|     :            |::|     | : : :
Db        118 L--------EDGLDERAARLSWLKTTLTDIAGG-------LPIIQGQNFGVLRYDESRHV 162

Qy        185 ----------FQSAERKG----EISQLKSMLTTTVAQERFAIQFKMENAKRCVTQLDRLS 230
                       |:|  :|     || |     || |           :    |:: | |:
Db        163 ENLALQNPKAMQTALNEGKINAEIFQRSVALTDT--------DYLRAKAQQIATSLNACK 214

Qy        231 ALVSTKCHSLGSQSTNFGFAKSLLTRVENALVHLSGIKLAPKAEAKTVEQEVAESSVSEG 290
              | ::     | :: :|     :|:|                          ::||  : :
Db        215 RLQASADRLFGHEAPSFASLADMLSRA----------------------SQLAEKLLKD- 251

Qy        291 ELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNN--------------MNRD 335
                      : :|  |   |:||  :    | || | |   |                : ||
Db        252 --------RGVELHPAIPEEAQPQAM---AANLHEAGEPMNATPPTTAPAALRTTPLTRD 300

Qy        336 LAFHLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQN-----GDALST 390
              || :|  :: ||:|||| ||: :|:|:||:|| : |   | ::: : |       | | |
Db        301 EAFTMLAGIAQYFKQSEPQSPVPYLIERAIKWGNMPLEGWLSDVIKDNNVVDNIRDVLGT 360


RESULT 8
Q9UPS3
ID   Q9UPS3      PRELIMINARY;      PRT;  1454 AA.
AC   Q9UPS3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein KIAA1079.
GN   KIAA1079.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99397452; PubMed=10470851;
RA   Kikuno R., Nagase T., Ishikawa K., Hirosawa M., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIV.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:197-205(1999).
DR   EMBL; AB029002; BAA83031.1; -.
DR   HSSP; P12931; 1FMK.
```

```
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR008266; Tyr_pkinase_AS.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Hypothetical protein; ATP-binding; Kinase; Transferase;
KW   Tyrosine-protein kinase.
SQ   SEQUENCE   1454 AA;  159543 MW;  86B23F1C453AD559 CRC64;

  Query Match             6.5%;  Score 152.5;  DB 4;  Length 1454;
  Best Local Similarity   21.4%;  Pred. No. 0.11;
  Matches  122;  Conservative   75;  Mismatches  189;  Indels  183;  Gaps   25;

Qy          3 LSKHQIEQLSKPLSD---DSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERD 59
              | | |:||   | ||   : :   :|  ||   ||   : :   | ||  ||  |  | |
Db        370 LPKPQLEQ---PYSDRWYEVLQFCWLSPEK---RPAAEDVHRLLTYLRLQSQRDS--EVD 421

Qy         60 ALQEACLNKWKILSDSLYEQFSKTTRDIELISWFVAAQF------LLDTTLESAANSLEW 113
                |:     |  |  :   : |       ::  |   :       :|  |  |   | |:
Db        422 FEQQ-----WNALKPNTNSRDSSNNAAFPILDHFARDRLGREMEEVLTVTETSQGLSFEY 476

Qy        114 LADLSE-KHWD-----HLN----------PVLPVETLKSDDDKGKEREQADAKVKAFFQL 157
              : : ::  |:|     ||:          ||   |:  ||   ||:
Db        477 VWEAAKHDHFDERSRGHLDEGLSYTSIFYPVEVFESSLSDPGPGKQ-------------- 522

Qy        158 VGDSEESSILYAPVLQLPLVGEVTFFD----------FQSAERKG---EISQLKSMLTT 203
                || :   |  |       | |  ||             |  |: |   |:    ::|||
Db        523 -DDSGQDVPLRVP-------GVVPVFDAHNLSVGSDYYIQLEEKSGSNLELDYPPALLTT 574

Qy        204 TVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVH 263
                      |:| :|   :| :|:|| | :   |   ||:  | :|     :::|
Db        575 ----------DMDNPERTGPELSQLTALRSVE---LEESSTDEDFFQSSTDPKDSSLPG 620

Qy        264 LSGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQ----TVSQHLH 319
              :   |::   :  :| :|     :||||   |  : : :   ||      |:|  |
Db        621 DLHVTSGPESPFNNIFNDVDKSE----DLPSHQ--KIFDLMELNGVQADFKPATLSSSLD 674

Qy        320 --------------------------AGNLSELGNLNNMNRDLAFHLLREVSDYFRQSE 352
                                       : ||   | : :|      ::|:|: |
Db        675 NPKESVITGHFEKEKPRKIFDSEPLCLSDNLMHQDNFDPLN-------VQELSENF---- 723

Qy        353 PHSPISFLLEKAIRWGYLSLPELLREMMSE-QNGDALSTIFNAAGLNHLD---------- 401
                || || :  | ||  | : :: :| :|     :   :  | ||
Db        724 -----LFLQEKNLLKGSLSSKEHINDLQTELKNAGFTEAMLETSCRNSLDTELQFAENKP 778

Qy        402 ----------------QVLL-----------PEVSTPTVGIESPQTPQAKPSVSDPRSV 433
                             |:|             ||| |    |: :|:|:  |  | |
Db        779 GMSLLQENVSTKGDDTDVMLTGDTLSTSLQSSPEVQVPPTSFETEETPRRVPPDSLPTQG 838

Qy        434 EEHVSQTSPVDTQSKQDQKPQSSATSALS 462
              |   :| : :|    :|   |  :  |::
Db        839 E---TQPTCLDVIVPEDCLHQDISPDAVT 864


RESULT 9
Q8IWU2
ID   Q8IWU2      PRELIMINARY;      PRT;  1503 AA.
AC   Q8IWU2;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   KPI-2 protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=12393858;
RA   Wang H., Brautigan D.L.;
RT   "A Novel Transmembrane Ser/Thr Kinase Complexes with Protein
RT   Phosphatase-1 and Inhibitor-2.";
RL   J. Biol. Chem. 277:49605-49612(2).
DR   EMBL; AY130988; AAN08717.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR   GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR008266; Tyr_pkinase_AS.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
SQ   SEQUENCE   1503 AA;  164925 MW;  EE1BCD66145CD226 CRC64;

  Query Match             6.5%;  Score 152.5;  DB 4;  Length 1503;
  Best Local Similarity   21.4%;  Pred. No. 0.12;
  Matches  122;  Conservative   75;  Mismatches  189;  Indels  183;  Gaps   25;

Qy          3 LSKHQIEQLSKPLSD---DSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERD 59
              | | |:||   | ||   : :   :|  ||   ||   : :   | ||  ||  |  | |
Db        370 LPKPQLEQ---PYSDRWYEVLQFCWLSPEK---RPAAEDVHRLLTYLRLQSQRDS--EVD 421

Qy         60 ALQEACLNKWKILSDSLYEQFSKTTRDIELISWFVAAQF------LLDTTLESAANSLEW 113
                |:     |  |  :   : |       ::  |   :       :|  |  |   | |:
Db        422 FEQQ-----WNALKPNTNSRDSSNNAAFPILDHFARDRLGREMEEVLTVTETSQGLSFEY 476

Qy        114 LADLSE-KHWD-----HLN----------PVLPVETLKSDDDKGKEREQADAKVKAFFQL 157
              : : ::  |:|     ||:          ||   |:  ||   ||:
Db        477 VWEAAKHDHFDERSRGHLDEGLSYTSIFYPVEVFESSLSDPGPGKQ-------------- 522

Qy        158 VGDSEESSILYAPVLQLPLVGEVTFFD----------FQSAERKG---EISQLKSMLTT 203
                || :   |  |       | |  ||             |  |: |   |:    ::|||
Db        523 -DDSGQDVPLRVP-------GVVPVFDAHNLSVGSDYYIQLEEKSGSNLELDYPPALLTT 574

Qy        204 TVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLLTRVENALVH 263
                      |:| :|   :| :|:|| | :   |   ||:  | :|     :::|
Db        575 ----------DMDNPERTGPELSQLTALRSVE---LEESSTDEDFFQSSTDPKDSSLPG 620

Qy        264 LSGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTKHIERIPMASEQAQ----TVSQHLH 319
              :   |::   :  :| :|     :||||   |  : : :   ||      |:|  |
Db        621 DLHVTSGPESPFNNIFNDVDKSE----DLPSHQ--KIFDLMELNGVQADFKPATLSSSLD 674

Qy        320 --------------------------AGNLSELGNLNNMNRDLAFHLLREVSDYFRQSE 352
                                       : ||   | : :|      ::|:|: |
Db        675 NPKESVITGHFEKEKPRKIFDSEPLCLSDNLMHQDNFDPLN-------VQELSENF---- 723

Qy        353 PHSPISFLLEKAIRWGYLSLPELLREMMSE-QNGDALSTIFNAAGLNHLD---------- 401
                || || :  | ||  | : :: :| :|     :   :  | ||
Db        724 -----LFLQEKNLLKGSLSSKEHINDLQTELKNAGFTEAMLETSCRNSLDTELQFAENKP 778

Qy        402 ----------------QVLL-----------PEVSTPTVGIESPQTPQAKPSVSDPRSV 433
                             |:|             ||| |    |: :|:|:  |  | |
Db        779 GLSLLQENVSTKGDDTDVMLTGDTLSTSLQSSPEVQVPPTSFETEETPRRVPPDSLPTQG 838

Qy        434 EEHVSQTSPVDTQSKQDQKPQSSATSALS 462
              |   :| : :|    :|   |  :  |::
Db        839 E---TQPTCLDVIVPEDCLHQDISPDAVT 864


RESULT 10
```

```
Q86A36
ID   Q86A36      PRELIMINARY;      PRT;  2245 AA.
AC   Q86A36;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to Dictyostelium discoideum (Slime mold). myosin IJ heavy
DE   chain.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE=22092622; PubMed=12097910;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL   Nature 418:79-85(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Baumgart C.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC115607; AAO52027.1; -.
DR   GO; GO:0016459; C:myosin; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003774; F:motor activity; IEA.
DR   InterPro; IPR002710; DIL.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001609; myosin_head.
DR   InterPro; IPR004009; Myosin_N.
DR   Pfam; PF01843; DIL; 1.
DR   Pfam; PF00612; IQ; 6.
DR   Pfam; PF00063; myosin_head; 1.
DR   Pfam; PF02736; Myosin_N; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD003376; DIL; 1.
DR   ProDom; PD000355; myosin_head; 1.
DR   SMART; SM00015; IQ; 3.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 3.
SQ   SEQUENCE   2245 AA;  258682 MW;  4DA40956515A7685 CRC64;

  Query Match             6.4%;  Score 151;  DB 5;  Length 2245;
  Best Local Similarity   23.6%;  Pred. No. 0.26;
  Matches   81;  Conservative   57;  Mismatches  143;  Indels   62;  Gaps   13;

Qy         23 VYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACLNKWKILSDSLYEQFSK 82
              : |: |  : | :: : |  |  | :|    ::: :   |   |  |: ||:  : |
Db        964 IQLRAEARSLRTVQEQKNKLQEKLEELQWRLTSEAKRKQQ---LEDQKVKSDTTISELSS 1020

Qy         83 TTRDIEL-ISWFVAAQFLLDTTLESAANSL-EWLADLSEK-----HWDHLNPVLPVETLK 135
                 :|| :|        || : :|:   | | |: | |:      |   ||  |  :
Db       1021 NNDHLELQLSEIQLKYQELDKSNQSSQLQLSECLSKLEEQTQQLDHSSKLNKKLEKDLSD 1080

Qy        136 SDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQLPLVGEVTF----FDFQSAER- 190
               |  | : | :   :    |     ||| |   :   |     :: |    ||  | ||
Db       1081 QHDSIEKLQSQFNETEQQLQQFKQQSEELSSKLSKTTQ-----QLDFNKQEFDRLSQERD 1135

Qy        191 ------KGEISQLKSMLTTTVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQS 244
                   : || |||    :|: :: |::    :| :| | :|   : |:  :  ||| ||
Db       1136 TDNTNNQLEIQQLKK-ANSTLEEDYFSLSGIRDNLERQVLELRDENQLIKERLDSLGQQS 1194

Qy        245 TNF--GFA------KSLLTRVENALVHLSGIKLAPKAEAK----------------TVE 279
               : |  | |      : |:       |: ||  ||  : |||                ::
Db       1195 SQFQSGAALEKQQLEQLVQEQSEQLIKLSSEKLGSEEEAKKQINQLELELTDHKSKLQIQ 1254

Qy        280 QEVAESSVS-----EGELPSHMDTK-----HIERIPMASEQAQ 312
               :: | |        :|:|  : | |      :|||  : :  :
Db       1255 LQLTEQSNEKIKKLKGKLEEYQDEKKQLQQELERIKQSKQSVE 1297
```

```
RESULT 11
Q86IW2
ID   Q86IW2      PRELIMINARY;      PRT;   959 AA.
AC   Q86IW2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE=22092622; PubMed=12097910;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL   Nature 418:79-85(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Baumgart C.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC116986; AAO51892.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   959 AA;  110909 MW;  151551A9D7D72335 CRC64;

  Query Match             6.1%;  Score 143.5;  DB 5;  Length 959;
  Best Local Similarity   19.7%;  Pred. No. 0.26;
  Matches   88;  Conservative   78;  Mismatches  173;  Indels  107;  Gaps   16;

Qy         41 VAQTALRKL-SQNPSADE--RDA--LQEACLNKWKILSD-------------SLYEQFS 81
              |: :| :||  ::  ||:  | |  || |    || : |               | |: |
Db        498 VSDSASKKLDKRDQEADKAIRQATDLQAALDKAWKDIQDLNKHHSIEIDRYNKQLQEERS 557

Qy         82 KTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDKG 141
              :   : :||:     :||          : :| |    ::|   :|      :| | :::
Db        558 RIKSEFQLIANNEKKEFL---------DKIEQLETFIQEHRSTINRNNEKQTWK-EEELN 607

Qy        142 KE-----REQADAKVKAFFQLVGDSEESSILYAPVLQLPLVGEVTFFDFQSAERKGEISQ 196
              ||      :   ||:::       |   |||   |    ||: ::     |  ||:
Db        608 KEIAHLQQRCRDAEIR------NDQLSSSI---PQATRPLLKQIESIQEQFNERQSTWET 658

Qy        197 LKSMLTTTVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNFGFAKSLLTR 256
              |:  |   : :||   :  : | :  :::|: |    : |   | |:|       |    :
Db        659 LEKQLNNQLREERLKAEHVVRNQQEILSELEEL----TLKFQLLESES------KIDKKK 708

Qy        257 VENALVHLSGIKLAPKAEAKTVEQEVAESSVSEGELPSHMDTKHI--ERIPMAS------ 308
              ::: :  | :|         ::   : ||  : ::  |:|  |  ||: :::
Db        709 LKSTITELQQLKSIQSDHLSKIQDLESSSSTLQSKIDQHLNTISILEERLTLSNKELKDL 768

Qy        309 ----------------------------EQAQTVSQHLHAGNLSELGNLNNMNRDLAFHL 340
                                         :| |   |  :   ::    | || | :
Db        769 EDKRQKEKELYEKETSNFLKKQQQLLLQQQLQQQQQQENQNQINNNNNNNNNNNNNGSS 828

Qy        341 LREVSDYFRQSEPHSP--------ISFLLEKAIRWGYLS-------LP---ELLREMMSE 382
              |   :|        :|        | | :      :|:       ||   | |:  :|:
Db        829 LSNSTDKINNGSISTPPQQQNKGLFSGLKKTPSNHDFLNHSVGNGLLPASLEYLQSSLSQ 888

Qy        383 QNGDALSTIFNAAGLNHLDQVLLPEV 408
              : |:|:|       |||  : |  |:
Db        889 KEGEAMSLQAQVQSLNHSRKKLEDEL 914


RESULT 12
O02239
ID   O02239      PRELIMINARY;      PRT;   884 AA.
AC   O02239; O02269;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
```

```
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   E01G4.6 protein.
GN   E01G4.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sims M.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Sims M.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z83223; CAB05719.1; -.
DR   EMBL; Z81538; CAB05719.1; JOINED.
DR   EMBL; Z81538; CAB04387.1; -.
DR   EMBL; Z83223; CAB04387.1; JOINED.
DR   PIR; T20405; T20405.
DR   WormPep; E01G4.6; CE15755.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0004623; F:phospholipase A2 activity; IEA.
DR   GO; GO:0016042; P:lipid catabolism; IEA.
DR   InterPro; IPR001211; PhospholipaseA2.
DR   PROSITE; PS00118; PA2_HIS; 1.
SQ   SEQUENCE   884 AA;  99603 MW;  73A1610965C9B30D CRC64;

  Query Match             6.1%;  Score 143;  DB 5;  Length 884;
  Best Local Similarity   20.7%;  Pred. No. 0.25;
  Matches  103;  Conservative   99;  Mismatches  165;  Indels  130;  Gaps   25;

Qy         27 LEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACLNKW----KILSDSLYEQFSK 82
Db        218 LRPDAYTEARITRRLALTSTATLSQR---------RERFLKKYSQIRQVVAQNLFGQQQR 268

Qy         83 TTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNPVLPVETLKSDDDKGK 142
Db        269 LTRPVEDLVTETSSKLAVEEAPEETTTQEETTTDASE-----------VTTTKAVEEATE 317

Qy        143 E-REQADAKVKAFFQLVGDSEESSILYAPVLQ-------------LPLVGEVTFFD-FQ 186
Db        318 EVTEEATEATEA---PVATTKESSEMHVNTIRNMIRSASEKDLSKYVTLISEGKFSELFE 374

Qy        187 SAERK------------GEISQLKSMLTTTVAQ-------ERFAIQFKMENAKRCVTQL 226
Db        375 LAEQKKLTLTSKFDEKLSSKMAKLKDLINEALSEKEKSGEIEQAMEKFEKPEKSELVAME 434

Qy        227 DRLSALVSTKCHSLGSQSTNFGFAKSLLTR----VENALVHLSGIKLAPKAEAKTVEQE- 281
Db        435 DKDTPAVFTISDSLKHKKAEAKLAHTIVSRNVVEAENA---IEKEVVEPKAEEKKVKEED 491

Qy        282 ---VAESSVSE---GELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNM--N 333
Db        492 VKAVAEEKKEEKKPGKLP--MKIEKLEK-PVDTK-------------SENHELKKVLDD 534

Qy        334 RDLAFHLLREVSDYFRQSEPHSPISFLLEK---AIRWGYLSLPELLREMMSEQ---NGDA 387
Db        535 KERALLVESEIKNTAEETKPKVE-SFKSEETTVAID----DMPALEKEESAEKKETTGEP 589

Qy        388 LSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQTPQAKPSVSDPRSVEEHVSQTSPVDTQ- 446
Db        590 TTT---------------EAAVETTEASETPKPEAKPELLS--NLEDVLTLTTP-ETET 630

Qy        447 ---SKQDQKPQSSATSA 460
Db        631 IEGSGEREEPTTSAPAA 647


RESULT 13
Q9I750
ID   Q9I750      PRELIMINARY;      PRT;   344 AA.
AC   Q9I750;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein PA0082.
GN   PA0082.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004447; AAG03472.1; -.
DR   PIR; G83634; G83634.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   344 AA;  36679 MW;  3321BAC4C8DB025D CRC64;

  Query Match             6.0%;  Score 141.5;  DB 16;  Length 344;
  Best Local Similarity   21.1%;  Pred. No. 0.082;
  Matches   89;  Conservative   70;  Mismatches  144;  Indels  119;  Gaps   18;

Qy         15 LSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQ-NPSADERDALQEACLNKWKILS 73
Db         11 VSPDSPCGDDL------------EYDAAFLELERIAQGQPERQMGDAVLPAEPPEWPRVR 58

Qy         74 DSLYEQFSKTTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDHLNPVLPVET 133
Db         59 ALASELFGR-SKDLRVANLLLQSNVALD-GLDGLADGLLLVRELLGQYWDGVYPL----- 111

Qy        134 LKSDDDKGKEREQADAKVKAFFQLVGDSEESSILYAPVLQL----PLVGEVTF--FDFQS 187
Db        112 LDADDD-----NDPTFRINALTGLVAE---------PLLQLVWAIPLVRSRAFGPVNLRA 157

Qy        188 AERKGEISQLKSMLTTTVAQERFAIQFKMENAKRCVTQLDRLSALVSTKCHSLGSQSTNF 247
Db        158 ALNAAGLQRFAS---ETLSPEQIAGAFADADA----------DALAATRRALDGAQEHAL 204

Qy        248 GFAKSLLTRVENA-------LVHL---------------SGIKLAPKAEAKTVEQEVAES 285
Db        205 AIESGVAERVGSAQGLDLGPLRQLLRQALQVFDLYGPQGAGESLAPGAEA-VADEQVGAA 263

Qy        286 SVSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNNMNRDLAFHLLREVS 345
Db        264 PVAAVAAPA----------PRAS-----------GEIA--------NREDVLRQLDRLL 293

Qy        346 DYFRQSEPHSPISFLLEKAIRWGYLSLPELLREMMSEQNGDALSTIFNAAGLNHLDQVLL 405
```

```
Db        294 EYYVRHEPSSPVPVLLKRAKTLVTADFAEIVRNLIPD------------GISQFETLRG 340

Qy        406 PE 407
              ||
Db        341 PE 342


RESULT 14
Q8MPV7
ID   Q8MPV7      PRELIMINARY;      PRT;  1534 AA.
AC   Q8MPV7;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein T23F2.2a.
GN   T23F2.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   Waterston R.;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Du Z.;
RT   "The sequence of C. elegans cosmid T23F2.";
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U39649; AAM69069.1; -.
DR   WormPep; T23F2.2a; CE31245.
KW   Hypothetical protein.
SQ   SEQUENCE   1534 AA;  177345 MW;  B82EF064300CCE52 CRC64;

  Query Match             6.0%;  Score 141;  DB 5;  Length 1534;
  Best Local Similarity   20.9%;  Pred. No. 0.75;
  Matches  115;  Conservative   90;  Mismatches  168;  Indels  178;  Gaps   27;

Qy          9 EQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACLNK 68
              :|||:  ||      :|  :::    |: |::   | |           |:   ::|  | |
Db        619 QQLSEHESD------FLSTQQT-LHDLKKEYSAKNTTL--------VDKFKEVEEILLAK 663

Qy         69 WKILSDSLYEQFSK----TTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDH 124
                ::: |:| :|        ||::          |  :  :    ||| :  ::||
Db        664 TELV-DALTKQLENIRKDQTRELS----------LKQSERDQYKKSLEEMTFIAEK---- 708

Qy        125 LNPVLPVETLKSDDDKGKEREQADAKVK---AFFQ-----LVGDS--------------- 161
                |:|  | |:   ||      :  |::|    :|:     |: ||
Db        709 -VPILEAEILQLSKDK----NEITARLKHDQEYFEDELAKLLNDSMNIKKERDDYLTEHI 763

Qy        162 --EESSILYAPVLQLPLVGEVTFFDFQSAER-KGEISQLKSMLTTTVAQERFAIQFKMEN 218
                 || |                        || | ||| ||  |      |       | |
Db        764 RANESMI---------------------ERLKLEISGLKKDLENQKMQAHLQ---KSEL 798

Qy        219 AKRCVTQLDRLSALVS--------TKCHSLGSQSTNFGFAK------------------ 251
                |: :: :| :| | |         :| ::  |   :  |
Db        799 EKKLLSSIDHVSQLQSRVNHSQRDVECQAIPRQINKYVGCKPNVKNKETIIEKGALFDEN 858

Qy        252 --------------SLLTRVENALVHL----------SGIKLAPKAEAKTVEQEVAESS 286
                           :| ::  || :           | :  :    ||  |  ||
Db        859 EERLRICKAELETTRRQVTVLQQKLVSIIQQQSSQKIKKRIAVVEDSNKNTVHTEDLESK 918

Qy        287 VSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGN------------LNNMNR 334
              :| ||    :|: :|||  : |  : |:  :    : :| |             ::| ::
Db        919 MKEVEL---KNTELMERID-SLEAERFVASSIEKSRIQKLVNEFDNLKQKLDNDMSNYSK 974

Qy        335 DLAF------HLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPE-LLREMMSEQN-GD 386
              :  :      :| :: |:  :| :| | |       ||::     || :  ||: ||| : ||
Db        975 EKQWLQWRISNLEKDNSELQKQIQPSS------EKSLE----SLNKGTLRKTMSEPDFGD 1024

Qy        387 ALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQT-PQAKPSVSDPRSVEEHVSQTSPVDT 445
               :||  : |  |      ::  |: | : ::|||   |     ::  ||  | |  |   :
Db       1025 DMSTEGDGASTNESADFMVESVTAPVLSLKSPQPFSQLADVLNLVRSDLEQV-LTEIEEP 1083

Qy        446 QSKQDQKPQSS 456
              :: : ::|| |
Db       1084 EAAKQEEPQMS 1094


RESULT 15
Q8MPV6
ID   Q8MPV6      PRELIMINARY;      PRT;  1538 AA.
AC   Q8MPV6;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein T23F2.2b.
GN   T23F2.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   Waterston R.;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Du Z.;
RT   "The sequence of C. elegans cosmid T23F2.";
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U39649; AAM69070.1; -.
DR   WormPep; T23F2.2b; CE31246.
KW   Hypothetical protein.
SQ   SEQUENCE   1538 AA;  177862 MW;  E6E8E4C16C941A85 CRC64;

  Query Match             6.0%;  Score 141;  DB 5;  Length 1538;
  Best Local Similarity   20.9%;  Pred. No. 0.76;
  Matches  115;  Conservative   90;  Mismatches  168;  Indels  178;  Gaps   27;

Qy          9 EQLSKPLSDDSICGVYLKLEKSAFRPLRNEFNVAQTALRKLSQNPSADERDALQEACLNK 68
              :|||:  ||      :|  :::    |: |::   | |           |:   ::|  | |
Db        619 QQLSEHESD------FLSTQQT-LHDLKKEYSAKNTTL--------VDKFKEVEEILLAK 663

Qy         69 WKILSDSLYEQFSK----TTRDIELISWFVAAQFLLDTTLESAANSLEWLADLSEKHWDH 124
                ::: |:| :|        ||::          |  :  :    ||| :  ::||
Db        664 TELV-DALTKQLENIRKDQTRELS----------LKQSERDQYKKSLEEMTFIAEK---- 708

Qy        125 LNPVLPVETLKSDDDKGKEREQADAKVK---AFFQ-----LVGDS--------------- 161
                |:|  | |:   ||      :  |::|    :|:     |: ||
Db        709 -VPILEAEILQLSKDK----NEITARLKHDQEYFEDELAKLLNDSMNIKKERDDYLTEHI 763

Qy        162 --EESSILYAPVLQLPLVGEVTFFDFQSAER-KGEISQLKSMLTTTVAQERFAIQFKMEN 218
                 || |                        || | ||| ||  |      |       | |
Db        764 RANESMI---------------------ERLKLEISGLKKDLENQKMQAHLQ---KSEL 798
```

```
Qy        219 AKRCVTQLDRLSALVS--------TKCHSLGSQSTNFGFAK------------------ 251
              |: :: :| :| | |         :| ::  |   :   |
Db        799 EKKLLSSIDHVSQLQSRVNHSQRDVECQAIPRQINKYVGCKPNVKNKETIIEKGALFDEN 858

Qy        252 ---------------SLLTRVENALVHL----------SGIKLAPKAEAKTVEQEVAESS 286
                          :| ::  || :             | :   :   ||  |  ||
Db        859 EERLRICKAELETTRRQVTVLQQKLVSIIQQQSSQKIKKRIAVVEDSNKNTVHTEDLESK 918

Qy        287 VSEGELPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGN------------LNNMNR 334
              : | ||    :|: :|||  : |  : |:  :    : :| |             ::| ::
Db        919 MKEVEL---KNTELMERID-SLEAERFVASSIEKSRIQKLVNEFDNLKQKLDNDMSNYSK 974

Qy        335 DLAF------HLLREVSDYFRQSEPHSPISFLLEKAIRWGYLSLPE-LLREMMSEQN-GD 386
              :  :      :| :: |:  :| :| |      ||::     || :  ||: ||| : ||
Db        975 EKQWLQWRISNLEKDNSELQKQIQPSS------EKSLE----SLNKGTLRKTMSEPDFGD 1024

Qy        387 ALSTIFNAAGLNHLDQVLLPEVSTPTVGIESPQT-PQAKPSVSDPRSVEEHVSQTSPVDT 445
              :||  : |  |     ::  |: | : ::|||   |    ::  ||  | |  |   :
Db       1025 DMSTEGDGASTNESADFMVESVTAPVLSLKSPQPFSQLADVLNLVRSDLEQV-LTEIEEP 1083

Qy        446 QSKQDQKPQSS 456
              :: : ::|| |
Db       1084 EAAKQEEPQMS 1094
```

Search completed: July 3, 2004, 02:46:15
Job time : 76 secs